



# Professionaliseren van lerend evalueren: lessen uit de praktijk van het Planbureau voor de Leefomgeving

## **Achtergrondpublicatie**

**Lisa Verwoerd**

**Hilde Brouwers**

**Evelien de Hoop**

**1-12-2022**





Planbureau voor de Leefomgeving



Athena Institute

Connecting science and society  
for a better tomorrow

## Colofon

### **Professionaliseren van lerend evalueren: lessen uit de praktijk van het Planbureau voor de Leefomgeving**

© PBL Planbureau voor de Leefomgeving

Den Haag, 2022

PBL-publicatienummer: 5047

## **Auteurs**

Lisa Verwoerd (PBL/VU), Hilde Brouwers (VU), Evelien de Hoop (VU)

## **Contact**

Lisa.verwoerd@pbl.nl

## **Met dank aan**

Alle respondenten die deelnamen aan het onderzoek waar deze publicatie op is gebaseerd, waaronder de onderzoekers betrokken bij de lerende evaluatie van het Natuurpact, het lerend kennisprogramma Regio Deals en de lerende evaluatie van het Interbestuurlijk Vitaal Platteland, en Jeannette Beck en Gusta Renes

De interne klankbordgroep Bas Arts, Femke Verwest en Eva Kunseler

De externe reflectanten Hanneke Drewes, Josefien de Bruin en Esther de Weger (allen RIVM) en Jitske van Popering en Michael Duijn (beiden GovernEUR, Erasmus Universiteit Rotterdam)

# Inhoud

Samenvatting en aanbevelingen	4
1. Introductie	10
2. Lerend evalueren: de theorie	12
3. Onderzoeksaanpak	15
4. Lerend evalueren: de PBL praktijk	18
5. Concluderende discussie en lessen	29
Referenties	<b>Error! Bookmark not defined.</b>

# SAMENVATTING EN AANBEVELINGEN

## Aanleiding en doel

Bestuurders, beleidsmakers en uitvoerders staan voor steeds complexere uitdagingen: weerbarstige problemen waarvoor de oplossing voor het één het volgende probleem is voor het ander, en waarover de meningen uiteen lopen over wat de gepaste weg voorwaarts is. In het licht hiervan wordt gezocht naar hoe kennisontwikkeling – zoals beleidsevaluatie – zo kan worden ingericht dat het recht doet aan deze complexiteit, en het beleid tijdig kan voorzien van relevante en bruikbare inzichten en zo het lerend vermogen van het beleid kan helpen vergroten. Een beleidsonderzoeksmethode die de laatste jaren aan populariteit wint is lerend evalueren.

Lerend evalueren is een participatieve evaluatiemethode waarbij beleidsonderzoekers en beleidsbetrokkenen tijdens alle fasen van de evaluatie gezamenlijk optrekken om die inzichten te ontsluiten die nodig zijn om de kwaliteit van het beleid te verbeteren. Het PBL pioniert met lerend evalueren sinds 2014. Dit gaat gepaard met successen, maar ook met de nodige obstakels. Projectteams zijn zoekende naar hoe lerend evalueren met kwaliteit en impact toe te passen. Die zoektocht vindt plaats in een organisatie- en beleidsomgeving die is gekenmerkt door een evaluatiepraktijk waarin meer traditionele evaluatiebenaderingen, gericht op doelbereik en met duidelijk gescheiden rollen voor onderzoekers en beleidsactoren, de norm zijn. Het vertrekpunt van deze publicatie is dan ook dat, om lerend evalueren met kwaliteit en impact toe te kunnen passen, deze methodiek dient te normaliseren – dat wil zeggen, inbedden in het bestaande evaluatierepertoire en ‘gewoon’ maken – en te professionaliseren bij beleidsonderzoeksorganisaties (of andere organisaties, zoals overheden, die lerend evalueren (willen) toepassen). Het normaliseerproces binnen zo’n organisatie staat daarbij natuurlijk niet los van zijn omgeving. Nationale en internationale (academische) ontwikkelingen binnen de (beleids)evaluatie community rekken het begrip van wat de norm is, op. En, vice versa, experimenten met nieuwe methodieken in de praktijk zijn voeding voor verdere innovaties in het veld. Deze publicatie richt zich op deze wisselwerking door van *binnen naar buiten* te kijken: door te kijken naar de uitdagingen die projectteams van lerende evaluaties ervaren met het toepassen en normaliseren van deze methodiek en de invloed van omgeving hierop.

Het doel van deze publicatie is het voorzien van lessen voor het normaliseren en professionaliseren van lerend evalueren bij beleidsonderzoeksorganisaties (of dus andere organisaties) die zich hiermee bezigen, of dat van plan zijn. Zij vormen eveneens de doelgroep. Hiertoe biedt de publicatie inzage in de praktijk van lerend evalueren bij het PBL, gebaseerd op casestudies van drie projecten (namelijk de lerende evaluatie van het Natuurpact, het lerend kennisprogramma Regio Deals en de lerende evaluatie van het Interbestuurlijk Vitaal Platteland). Daarnaast worden de uitdagingen die de respectievelijke projectteams tegenkomen met de methodiek beschreven.

## Bevindingen

Bij het PBL worden lerende evaluaties ontworpen als gefaseerde, participatieve kennisprogramma’s. Elke lerende evaluatie betreft maatwerk om zo goed aan te sluiten

op het voorliggende vraagstuk, de beleidscontext en de evaluatiebehoefte die daar leven. Dit geldt ook voor evaluatiecriteria die worden gehanteerd; zo richt de ene lerende evaluatie zich op doelbereik en de werking van beleidsstrategieën (en *governance*), en anderen op de mate waarin het beleid slaagt een transitie in gang te zetten. Inspiratie wordt daarvoor geput uit internationale evaluatiestudies naar het transformerend vermogen van beleid dat zich richt op duurzame ontwikkeling.

Voor het ontwerp en uitvoering maken onderzoekers bij PBL gebruik van een zestal 'richtinggevende principes' die diverse kenmerken van de methodiek beschrijven en die op orde dienen te zijn. Wat 'op orde' precies behelst, verschilt per project en hangt af van de respectievelijke beleidsdynamiek, -vragen, etc. De principes worden daarnaast doorontwikkeld naargelang de ervaring met de methodiek groeit. De richtinggevende principes zijn als volgt:

- i. Lerend evalueren dient om het lerend vermogen van beleid te vergroten en is geschikt voor complexe, interbestuurlijke beleidsprogramma's die doelzoekend zijn en gericht op een expliciete verander-(of transitie-)opgave
- ii. Lerend evalueren is een *ex durante* methodiek die parallel aan het beleidsproces wordt uitgevoerd en daarmee mee ontwikkelt. De methodiek is responsief naar beleidsontwikkelingen, en daardoor iteratief en emergent van aard
- iii. Lerend evalueren is maatwerk: het evaluatiecriterium en de inhoudelijke afbakening worden afgestemd op de behoeften in de beleidspraktijk. De geselecteerde evaluatiemethoden zijn gericht op reflectie en gezamenlijk leren, en zijn actiegericht
- iv. Lerend evalueren is participatief: relevante beleidsactoren spelen een actieve rol en ervaren eigenaarschap over het ontwerp en uitvoering. Er is daarbij oog voor inclusie en diversiteit ten behoeve van blinde vlekken, nieuwe oplossingen en legitimering van het lerende proces
- v. Lerend evalueren vergt een team met beleidsinhoudelijke en bestuurskundig expertise, én expertise met participatief en lerend onderzoek
- vi. Lerend evalueren vraagt zorg voor geloofwaardigheid, onafhankelijkheid en beheersbaarheid van het onderzoek, door transparant en open te zijn over werkzaamheden, keuzes en prioriteringen

De uitdagingen die projectteams tegenkomen zijn gecategoriseerd volgens het Normalisation Proces Theory kader van May en Finch (2009). Dit kader identificeert vier overlappende mechanismen die – op niet-volordelijke wijze – bijdragen aan het normaliseren van innovatieve werkwijzen. Deze mechanismen zijn *begripsvorming*, *draagvlak*, *toepassing* en *waardering*. De uitdagingen staan per NPT mechanisme weergegeven in Tabel I.

**Tabel I. De uitdagingen samengevat per NPT mechanisme.**

Begripsvorming	Draagvlak	Toepassing	Waardering
Uiteenlopende beelden binnen het projectteam over wat lerend evalueren is en doet	Beperkt draagvlak onder PBL-collega's	Interdisciplinaire kennisintegratie is stroperig proces	Waardering vindt vooral informeel plaats
Beperkt beschikbare informatie over de methode	Betrokkenheid en commitment van verzoekers niet vanzelfsprekend	Passende expertise en ervaring niet zomaar beschikbaar	Lessen landen lastig in vervolgprojecten
	Onduidelijk doel en strategie voor betrekken van regionale partijen	Sectoroverstijgend samenwerken niet vanzelfsprekend	Formele waardering vindt plaats na afloop van projecten, minder tijdens het project
		Iteratief en emergente werkwijze minder toegesneden op werkprogrammering	
		Verschillende snelheden van onderzoeks- en beleidspraktijk	

## Conclusies en lessen

De bevindingen laten zien dat een lerende evaluatiepraktijk zich met succes ontwikkelt bij het PBL. De methodiek is in toenemende mate een vanzelfsprekende optie voor beleidsonderzoek en lijkt daarmee steeds meer geaccepteerd als wetenschappelijk robuuste methode. Om dit proces van normaliseren en professionaliseren verder aan te moedigen, en zo de kwaliteit en impact waarmee lerend evalueren wordt toegepast te vergroten, identificeren we drie 'clusters' van lessen. Deze clusters zijn gericht op individuele projectteams, de 'community of practice' rondom lerend evalueren binnen de organisatie, en op het organisatieniveau.

De lessen die we trekken dienen eveneens ter inspiratie voor andere organisaties.

### ***I. Individuele projectteams: draag zorg voor het interne proces, en kwaliteits- en omgevingsbewustzijn***

We constateren dat individuele projectteams bij het PBL nog bewuster en meer intentioneel kunnen investeren in en sturen op de kwaliteit en impact van hun lerende evaluatie.

- *Het interne proces.* Voorafgaand aan een lerende evaluatie is een gesprek nodig over de beelden die er spelen over wat lerend evalueren is en doet, wat de motivatie is voor een lerende evaluatieaanpak gegeven het voorliggende verzoek, en hoe wordt gedacht over PBL's rol. In aanvulling hierop is het wenselijk om al aan de voorkant van het project een plan voor interdisciplinaire samenwerking te maken. Kennisintegratie gaat niet vanzelf, maar vergt structurele tijdsinvestering en een ingericht proces.

- *Kwaliteitsbewustzijn.* Het kwaliteitsbewustzijn kan worden vergroot door meer actief de richtinggevende principes voor lerend evalueren te hanteren en hier regelmatig – tijdens de lerende evaluatie – op te reflecteren. Investeer aanvullend in een projectevaluatie achteraf, specifiek gericht op de werking van de methodiek en op het beschikbaar stellen van deze lessen.
- *Omgevingsbewustzijn.* Vanuit individuele projecten kan meer worden gestuurd op impact door het omgevingsbewustzijn te versterken. Ook hier is regelmatige reflectie noodzakelijk. We achten het daarbij behulpzaam als wordt ingezet op de ontwikkeling van een impactstrategie waarin de rol en aanpak van het PBL in relatie tot de verschillende betrokken partijen duidelijk wordt gemaakt.

## **II. Community of Practice: investeer in lerend vermogen rondom lerend evalueren**

Er is binnen PBL, hoewel klein, een informele *community of practice (CoP)* rondom lerend evalueren. Hier ligt een kans voor de community om het lerend vermogen rondom lerend evalueren te versterken. Hiervoor zien we verschillende mogelijke aandachtspunten.

- *Verstevig de Community of Practice.* CoP's zijn geïdentificeerd als een effectieve wijze van kennismanagement en leerprocessen binnen organisaties. Door met regelmaat lessen uit te wisselen, bijvoorbeeld van methodologische aard maar ook omtrent bepaalde uitdagingen, kan zo'n CoP meer worden verstevigd. Dit betekent niet gelijk dat er een formele CoP moet worden georganiseerd met regelmatige overleggen, maar wel dat ervaringsdeskundigen worden aangemoedigd meer actief met elkaar de uitwisseling op te zoeken. Dit kan ook de vorm hebben van bijvoorbeeld intervisie. Wie onderdeel uitmaakt van de CoP is daarbij dynamisch. We achten het kansrijk om bijvoorbeeld ook eens beleidsonderzoekers uit te nodigen met ervaring met andere vormen van participatief onderzoek dan evaluatie specifiek, en mogelijk ervaringsdeskundigen van andere organisaties.
- *Werk aan een beschikbare kennisbasis.* Een toegankelijke kennisbasis met daarin informatie over de achtergrond en theorie van lerend evalueren en die wordt aangevuld naarmate er meer lessen worden opgedaan (bijvoorbeeld middels individuele projectevaluaties), kan een waardevolle bron voor lerend vermogen zijn. Denk bijvoorbeeld aan handreikingen, conceptuele kaders, toolboxes, etc. Deze kennis dient vindbaar en toegankelijk te zijn.
- *Maak keuzes en stel randvoorwaarden.* Het toenemend aantal verzoeken voor lerende evaluaties en de beperkte capaciteit bij het PBL vraagt om keuzes te maken en om prioritering. Niet alles kan. Daarbij laat deze studie een diversiteit aan invullingen van lerend evalueren zien, gericht op verschillende type vraagstukken. Een overweging is op zijn plaats of het effectiever zou zijn om uitsluitend te richten op een bepaald type vraagstuk en daaraan randvoorwaarden te stellen die de impact van een lerende evaluatie vergroten. Bijvoorbeeld door alleen verzoeken voor lerende evaluaties van transitiebeleid te honoreren, waarvoor een expliciete leerambitie is uitgesproken.

### **III.      *Organisatieniveau: wees responsief naar de toepassing en doorontwikkeling van lerend evalueren***

Hoewel lerend evalueren steeds meer verankerd raakt in de organisatie, observeren we dat verschillende factoren die op organisatieniveau spelen de succesvolle toepassing van lerende evaluaties bemoeilijken. We identificeren hiervoor een aantal kansen.

- *Organisatiestructuren.* Op enkele vlakken sluiten organisatiestructuren niet goed aan om de praktijk van lerend evalueren goed in te kunnen bedden. Zo past de wijze van werkprogrammering bijvoorbeeld niet goed bij het iteratieve en emergente karakter van lerend evalueren, en is sectoroverstijgend samenwerking geen vanzelfsprekendheid. Een meer flexibelere werkprogrammering en meer incentive voor samenwerking tussen sectoren (op beide vlakken zijn bij PBL reeds stappen gezet) zou de kwaliteit van lerend evalueren bevorderen omdat dan het iteratieve proces meer ruimte krijgt, en interdisciplinaire samenwerking wordt vergemakkelijkt.
- *Borging van kennisbasis.* De kennisbasis en poule van beleidsonderzoekers die PBL in huis heeft voor het uitvoeren van lerende evaluaties is klein. Expertise met transdisciplinaire, deliberatieve en lerende onderzoeksmethodieken is beperkt beschikbaar en ervaring die hiermee wordt opgedaan dreigt verloren te gaan vanwege tijdelijke aanstellingen. Door tijdelijke aanstellingen te bestendigen in vaste wordt de opgedane ervaring behouden (ook hier zijn relevante ontwikkelingen gaande). Aanvullend zou de kennisbasis kunnen worden versterkt door trainingen en workshops voor participatieve en deliberatieve onderzoeksmethoden aan te bieden aan beleidsonderzoekers.

Het onderzoek waar deze publicatie op is gebaseerd onderschrijft de complexe wisselwerking tussen de evaluatiepraktijk van een beleidsonderzoeksorganisatie, en die van zijn omgeving. De toenemende vraag naar evaluatieonderzoek dat zich richt op het leren binnen (en tussen) beleidsprogramma's sterkt het normaliseerproces. Andersom dragen succesverhalen van lerende evaluaties bij aan deze vraag (en dus het normaliseren van de methode binnen de beleidspraktijk). Dit is belangrijk, omdat het succes van een lerende evaluatie afhangt van de mate waarin deze evaluatiebenadering voet aan de grond heeft in de beleidspraktijk en onder de betrokken deelnemers. Als de bereidheid tot leren bij hen beperkt is en/of als de structuren die nodig zijn om dit leren te doen plaatsvinden afwezig zijn, wordt de kwaliteit en impact van lerend evalueren serieus beperkt. Enkel met een lerende houding en structuren die dit faciliteren wordt het mogelijk om lerend evalueren uit te voeren op een wijze die het impactpotentieel volledig benut.

Diverse studies laten zien dat hoewel leren in zwang lijkt, dit bij de overheid nog niet altijd navolging krijgt noch tot de verwachte successen leidt. Leren en lerend evalueren is niet gewoon. Naast een lineaire resultaatgerichtheid en geïnstitutionaliseerde verantwoordingsmechanismen hangt dit in ieder geval ten dele samen met wat geldt als de norm voor beleidsevaluatie bij de evaluatieafdelingen van o.a. ministeries. Zolang in de evaluatie community uitsluitend doeltreffendheid, doelmatigheid en dus het één-op-één aantonen van causale relaties als dé na te streven gouden standaarden voor evaluatie worden gehanteerd, zal het een uitdaging blijven om met lerend evalueren serieus te worden genomen. Lerend evalueren wordt dan gezien als een



vrijblijvende, 'laagwaardige' evaluatiemethode, in plaats van als een actiegerichte en impactvolle methodiek die vooruitgang weet te boeken in het gezicht van weerbarstige opgaven. Niettemin wekken de diverse experimenten met lerend evalueren die plaatsvinden binnen de Rijksoverheid de suggestie dat lerend evalueren een voet tussen deur heeft, en schetsen een gunstig beeld wat dat betreft.

We eindigen deze publicatie dan ook met het onderstrepen van het belang dat lerend evalueren niet alleen normaliseert en professionaliseert bij beleidsonderzoeksorganisaties, maar ook bij de overheden wiens beleid wordt onderzocht en de aanverwante evaluatie community. Ook dáár dient lerend evalueren te normaliseren en professionaliseren als serieuze evaluatiemethodiek voor complexe (transitie-)vraagstukken.

# 1. Introductie

## 1.1 Aanleiding

Bestuurders, beleidsmakers en uitvoerders staan voor steeds complexere uitdagingen. Het betreffen veelal 'wicked problems': weerbarstige problemen waarvoor de oplossing voor het één het volgende probleem is voor het ander, en waarover de meningen uiteen lopen over wat de gepaste weg voorwaarts is (Rittel & Webber, 1973). Van het fysiek-ruimtelijke domein (klimaatverandering, de energietransitie), het sociaaleconomische (de kloof tussen rijk en arm, de toeslagenaffaire) tot het sociale domein (toegang tot zorg, jeugdzorg) vragen deze uitdagingen om nieuwe vormen van samenwerking tussen overheden onderling, en met partijen uit de samenleving (Kunseler et al., 2020). De complexiteit van deze vraagstukken maakt dat vooraf niet volledig vast valt te stellen wat zal werken. Tegelijkertijd laat de urgentie niet toe om af te wachten tot daar volledig duidelijkheid over ontstaat – voor zover dat überhaupt mogelijk is, gegeven de grillige omgeving waarbinnen deze maatschappelijke opgaven spelen (Weterings & Verwoerd, 2020). Gepaard met nieuwe wijzen waarop beleidsontwikkeling en -uitvoering worden georganiseerd, is daarom een zoektocht naar hoe kennisontwikkeling – zoals beleidsevaluatie – voor de beleidspraktijk zo kan worden ingericht dat het recht doet aan deze complexiteit en het beleid tijdig kan voorzien van relevante en bruikbare inzichten over 'wat werkt en waarom' (Houppermans, 2018; van der Steen et al., 2018). De uitdagingen zoals hierboven beschreven behoeven namelijk een wijze van evalueren die is gericht op gedurende het beleid leren over de oorzaken van het probleem en welke aanpak daarvoor werkt, en die dus parallel loopt aan het beleidsproces (ex durante; van der Steen et al., 2018). Een beleidsonderzoeksmethode die in dit licht de laatste jaren aan populariteit wint, is lerend evalueren.

Lerend evalueren is een participatieve evaluatiemethode waarbij beleidsonderzoekers en beleidsbetrokkenen tijdens alle fasen van de evaluatie gezamenlijk optrekken om die inzichten te ontsluiten die nodig zijn om de kwaliteit van het beleid te verbeteren (L. Verwoerd et al., 2020). Een snelle Google-search laat zien dat de methode aan terrein wint bij zowel overheden (van gemeenten, provincies en vakdepartementen aan toe) als (beleids)onderzoekorganisaties (van consultancies tot regionale planbureaus, en nationale instituten zoals het RIVM). Zo ook bij het Planbureau voor de Leefomgeving (PBL).

Het PBL pioniert met lerend evalueren sinds 2014. In de PBL Visie 2025 (2017) werd gesteld dat, gegeven de complexiteit en onzekerheid omtrent leefomgeving vraagstukken en hun multi-level multi-actor governance karakter, innovatie richting meer transdisciplinaire en deliberatieve onderzoeksmethoden nodig en urgent is (zie ook Kunseler & Verwoerd, 2019). Deze in eerste instantie interne constatering vindt zijn weerklank in PBL's beleidsomgeving. In toenemende mate wordt het PBL gevraagd om ex durante beleidsonderzoek uit te voeren naar complexe, interbestuurlijke beleidsprogramma's – zoals het Programma Aardgasvrije Wijken en, meer recentelijk, het Klimaatbeleid – gericht op vraagstukken uit de leefomgeving, om zo al tijdens de uitvoering van dergelijke programma's de kwaliteit van besluitvorming te vergroten. Niettemin spelen er binnen PBL uitdagingen voor het uitvoeren van lerende evaluaties en het professionaliseren van deze methodiek. Uit de PBL-praktijk klinken geluiden van projectteams die zoekende zijn naar hoe lerend evalueren met kwaliteit en

impact toe te passen. Die zoektocht vindt plaats in wisselwerking met een omgeving (Nb. zowel de organisatie zelf, als de beleidsomgeving en aanverwante evaluatie *community*) waarin meer 'traditionele' benaderingen van evaluatie – gericht op doelbereik en verantwoording – meer gewoon zijn. Er spelen (meer of minder impliciete) verschillende beelden over wat evaluatie 'is' en hoe je dit 'goed doet' (Kunseler, 2017) die de ruimte bepalen voor wat voor benadering legitiem is (Verwoerd, 2022). Een recente studie toont aan dat vergelijkbare uitdagingen ook spelen bij andere organisaties die zich bezigen met lerend evalueren, zoals het RIVM, het Fries Sociaal Planbureau en bijvoorbeeld het ministerie van Volksgezondheid, Welzijn en Sport (uitgevoerd in opdracht van het PBL, (Bloemert, 2022; zie ook Regioplan, 2022).

Het vertrekpunt van deze publicatie is dan ook dat, om lerend evalueren met kwaliteit en impact toe te kunnen passen, deze methodiek dient te *normaliseren* – dat wil zeggen, inbedden in het bestaande evaluatierepertoire en 'gewoon' maken – en te professionaliseren bij beleidsonderzoeksorganisaties (of andere organisaties, zoals overheden, die lerend evalueren (willen) toepassen). Het normaliseerproces binnen zo'n organisatie staat daarbij natuurlijk niet los van zijn omgeving. Nationale en internationale (academische) ontwikkelingen binnen de (beleids)evaluatie community rekken (of verengen) het begrip van wat de norm is, op. En, vice versa, experimenten met nieuwe methodieken in de praktijk zijn voeding voor verdere innovaties in het veld. Deze publicatie richt zich op deze wisselwerking door van *binnen naar buiten* te kijken: door te kijken naar de uitdagingen die projectteams van lerende evaluaties ervaren met het toepassen en normaliseren van deze methodiek en de invloed van omgeving hierop. Om het proces van normaliseren en professionaliseren te ondersteunen biedt deze publicatie inzage in de praktijk van lerend evalueren bij het PBL – als paradigmatische casus (Flyvbjerg, 2006) – om daar lessen uit te trekken.

## 1.2 Doel en vraagstelling

Het doel van deze publicatie is het voorzien van lessen voor het normaliseren en professionaliseren van lerend evalueren bij beleidsonderzoeksorganisaties (of dus andere organisaties) die zich hiermee bezigen, of dat van plan zijn. Zij vormen eveneens de doelgroep. Om dit doel te realiseren bieden we inzage in hoe lerend evalueren als praktijk invulling krijgt bij het PBL. Hiervoor nemen we drie lerende evaluatie projecten – de lerende evaluatie van het Natuurpact, het lerend kennisprogramma Regio Deals en de lerende evaluatie van het Interbestuurlijk Vitaal Platteland – onder de loep. We beschrijven de gemeenschappelijke uitdagingen die de betrokken projectteams tegenkwamen met betrekking tot het toepassen van lerend evalueren, en hoe deze kunnen worden verklaard en vervolgens mogelijk het hoofd geboden.

Het onderzoek dat de grondslag biedt voor dit artikel had als doel de praktijk bij het PBL in kaart brengen en het normaliseren en professionaliseren van lerend evalueren bij de organisatie te ondersteunen. De lessen uit die studie worden in de huidige notitie veralgemeeniseerd zodat ze ook van toepassing zijn voor andere beleidsonderzoeksorganisaties.

De hoofdvraag luidt: *Wat zijn lessen voor het normaliseren en professionaliseren van lerend evalueren bij beleidsonderzoeksorganisaties, zoals het PBL?* Hiervoor beantwoorden we de volgende deelvragen:

- i. Hoe ziet de praktijk van lerend evalueren bij het PBL er op dit moment uit?
- ii. Tegen welke gemeenschappelijke uitdagingen lopen projectteams van lerende evaluaties aan met het toepassen van deze methode en hoe kunnen deze worden verklaard?

### 1.3 Leeswijzer

Paragraaf 2 schetst theoretische achtergrond van deze studie. Het beschrijft beknopt de ontwikkelingen die hebben geleid tot de opkomst van lerend evalueren. Daarnaast wordt het kader dat is gebruikt om te reflecteren op uitdagingen met het normaliseren en professionaliseren van deze methodiek bij het PBL beschreven. Paragraaf 3 beschrijft de onderzoeksaanpak, gevolgd door de resultaten in 4. Paragraaf 5 betreft de concluderende discussie en beschrijft tenslotte de lessen die wij trekken.

## 2. Lerend evalueren: de theorie

In deze theoretische achtergrond volgt een beknopte beschrijving van de aanleiding voor de opkomst van lerend evalueren en de 'evaluatieschool' waar deze methodiek zijn grondslag in vindt. Daarna wordt het theoretisch kader beschreven dat in deze studie is gebruikt ter analyse van de toepassing en het normaliseren van lerend evalueren bij het PBL als nieuwe (onderzoeks)praktijk, namelijk *Normalisation Process Theory*, NPT.

Voorop gesteld, lerend evalueren is geen gestandaardiseerde evaluatiemethode (zoals bijvoorbeeld Michael Patton's *utilization-focused evaluation*, 2000; of *realist evaluation* van Pawson & Tilley, 1997). Dat wil zeggen, hoewel de naam momenteel veelvuldig wordt gebruikt is het geen gestandaardiseerde methodiek waarvoor (min of meer) klip en klare handreikingen of kaders beschikbaar zijn, en waarvoor bijvoorbeeld trainingen te volgen zijn. Tegelijkertijd is het ook niet zo dat '*anything goes*' – lerend evalueren valt in ieder geval onder de reflexieve evaluatieschool en de *participatory evaluation*-tak, waaronder een aantal herkenbare principes liggen.

### 2.1 Roep naar reflexieve en deliberatieve evaluatiemethoden

De opkomst van lerend evalueren kan worden geplaatst in een bredere beweging naar meer reflexieve en deliberatieve evaluatiemethoden. Er is de afgelopen decennia meer herkenning gekomen voor de beperkingen van wetenschap als ultiem antwoord op complexe duurzaamheidsvraagstukken, en voor het belang van het integreren van kennis uit de praktijk om besluitvormingsprocessen te voeden (Sarewitz, 2004). Kennis uit een ivoren toren loopt immers het risico weinig recht te doen aan de belevingen en waarden die leven in de samenleving, omdat het een eenzijdig beeld geeft van een vraagstuk. Zo'n vraagstuk gaat bovendien gepaard gaat met veel aannames en onzekerheden die, aldus de retoriek, dienen te worden erkend, en openbaar en transparant te worden gemaakt. Om deze redenen wordt gezocht naar methoden waarbij het proces van kennisproductie gezamenlijk wordt vormgegeven door wetenschappers, beleidsmakers en andere relevante stakeholders uit de maatschappij. De gedachte is daarbij dat dit proces van kennis coproductie kennis ontwikkelt die niet alleen wetenschappelijk, maar ook sociaal robuust is (en daarmee recht doet aan de meervoudige perspectieven uit de samenleving), en kan

bijdragen aan maatschappelijke transitie en duurzame ontwikkeling (Regeer & Bunders, 2009; Scholz & Steiner, 2015).

**Tabel 2.1 Verschillen tussen de moderne en reflexieve evaluatiescholen.**

	Moderne school	Reflexieve school
<b>Functie van evalueren</b>	Evaluatie is een instrument om doeltreffendheid en doelmatigheid in kaart te brengen, ter verantwoording van het beleid en ter naleving van de afspraken	Evaluatie is een mechanisme om het lerend vermogen van beleid(sinterventies) ten behoeve van (systeem)transformatie te vergroten
<b>Perspectief op de kennis-beleidsinterface</b>	Relaties tussen kennis en beleid zijn technocratisch en bureaucratisch van aard; kennis en beleid zijn strikt gescheiden domeinen en dienen gescheiden te blijven	Relaties tussen kennis en beleid zijn open, transparant en deliberatief; kennis en beleid zijn inherent en op complexe wijze met elkaar verweven en die verwevenheid dient te worden (h)erkend
<b>Rol van evaluatoren</b>	Evaluatoren voorzien van onafhankelijke, empirisch bewezen en consistente feiten om zo evidence-based besluitvorming te voeden. Hiervoor zetten ze zich in om de domeinen kennis en beleid gescheiden te houden.	Evaluatoren faciliteren interactieve leerprocessen en systeemleren ten behoeve van (systeem)transformatie. Ze organiseren productieve interacties tussen kennis en beleid om sociaal robuuste antwoorden voor complexe maatschappelijke problemen te co-produceren.
<b>Methoden en tools</b>	Impact assessments, kosten/baten analyses, kosten-effectiviteit analyses, modelleren	Systeemanalyses, joint fact finding, reflexieve monitoring
<b>Epistemologische grondslag</b>	Positivism, empiricism	Sociaal constructivisme, pluralisme, pragmatisme
<b>Onderliggende waarden en principes</b>	Objectiviteit, neutraliteit, onpartijdigheid	Intersubjectiviteit, erkenning van meervoudigheid van perspectieven en onzekerheid, transparantie, bescheidenheid

De beweging naar meer reflexieve en deliberatieve onderzoeksmethoden wordt in de literatuur soms aangeduid als de opkomst van een 'nieuwe' school van (evaluatie)onderzoeksmethoden. Deze wordt al dan niet afgezet tegen de meer 'klassieke' of traditionele benadering, en verschillen tussen deze scholen worden geïdentificeerd in o.a. de onderliggende wetenschapsfilosofie (epistemologie), blik op de interactie tussen wetenschap, beleid en samenleving, en het doel van kennisproductie. Bijvoorbeeld 'moderne' vs. 'postmoderne' wetenschap (Funtowicz & Ravetz, 1993), 'technische' vs. 'deliberatieve' modellen (Owens et al., 2004), 'technocratische' vs. 'participatieve' methoden (Chouinard, 2013) of 'moderne' vs. 'reflexieve' logica (Kunseler & Vasileiadou, 2016), waarbij de eerste steeds een klassieke kijk op wetenschap – op afstand van het onderzoeksobject – belichaamt, en het tweede een reflexieve en deliberatieve kijk die de wetenschapper 'in' de samenleving plaatst.

Tabel 2.1 zet een aantal verschillen op een rij zoals geconceptualiseerd voor moderne en reflexieve evaluatiescholen (naar o.a. Kunseler & Vasileiadou, 2016; van Veen et al., 2016; Verwoerd et al., 2021). Lerend evalueren is een methodiek die zijn oorsprong vindt in de reflexieve school.

Hoewel in de praktijk zulke scholen zelden in een heel duidelijk afgebakende vorm voorkomen, helpt zo'n differentiatie om grip te krijgen op verschillende bestaande paradigma's over wat 'goed' evaluatieonderzoek is en doet. Wat vanuit de ene logica namelijk kan worden gezien als een gouden standaard voor evalueren kan vanuit de ander begrepen worden als een ernstige tekortkoming. Bijvoorbeeld, de afstand van een evaluator tot het evaluatieobject is vanuit het modern perspectief heilig om tot een onafhankelijk oordeel te kunnen komen. Echter, volgens een reflexieve blik zijn participatie en deliberatie cruciaal zijn voor een goede aansluiting op de (beleids)praktijk, en de relevantie en bruikbaarheid van het onderzoek (wat overigens niet betekent dat onafhankelijkheid onbelangrijk wordt gevonden bij reflexieve evaluatie). Bovendien, zo laten diverse studies zien, maakt dat de moderne school de historische oudere en daardoor meer ingebedde manier van (denken over) evalueren is, dat er in de praktijk niet altijd ruimte en begrip is voor de participatieve en lerende manier (e.g. Chouinard, 2013). Om die reden loopt deze school van evalueren in de toepassing nogal eens tegen barrières aan.

Om meer inzicht en grip te krijgen in hoe methodieken zoals lerend evalueren voet aan de grond kunnen krijgen bij organisaties die meer gewoon zijn met traditionele wijzen van evalueren, is het zinvol om te kijken hoe een praktijk als lerend evalueren bij het PBL gestalte krijgt en waar projectteams zoal tegenaan lopen met het toepassen en inbedden van de methodiek in het evaluatierepertoire van hun organisatie. Puttend uit de implementatie- en organisatiewetenschappen, noemen we dit proces normaliseren. In de volgende paragraaf conceptualiseren we hoe dit proces van normaliseren kan worden bestudeerd.

## 2.2 Normaliseren en professionaliseren in beeld

De toevoeging van lerend evalueren aan PBL's onderzoeksrepertoire kan worden gezien als een innovatie van PBL's onderzoekspraktijk<sup>1</sup>. Om te begrijpen waarom sommige innovaties wel of niet houvast vinden – normaliseren – binnen een organisatie wordt aanbevolen het dynamische proces van implementatie en vervolgens normaliseren volledig in beschouwing te nemen (Fullan & Pomfret, 1977). Een kader dat zich hier goed voor leent is het *Normalisation Process Theory framework* (NPT, May & Finch, 2009).

NPT is bedoeld om te kunnen begrijpen waarom innovaties wel of niet verworden tot standaard praktijk. Het gaat uit van vier gelijktijdige en samenhangende 'mechanismen' die de handelingen en acties omvatten van practitioners die zij dienen te ondernemen om het normalisatieproces te doen slagen. De theorie vertrekt vanuit een *practice perspective*: niet individuele meningen of intenties staan centraal, maar juist

---

<sup>1</sup> Waarbij we overigens erkennen dat deze innovatie voortbouwt op een langere geschiedenis van toepassing van participatieve en deliberatieve onderzoeksmethoden bij het planbureau, zie bijvoorbeeld Kunseler & Vasileiadou (2016), Kunseler (2016) en Petersen et al. (2011).

de waarneembare, collectieve en georganiseerde handelingen die mensen intentioneel nemen (Nicolini, 2012). De vier mechanismen beslaan handelingen en acties ter *begripsvorming* over; ontwikkelen van *draagvlak* voor; en *toepassing* en *waardering* van een innovatie (tabel 2). Belangrijk is dat deze mechanismen niet los van elkaar of chronologisch in de tijd moeten worden begrepen: ze hebben alle betrekking op elkaar, en acties op het ene mechanisme zullen van betekenis zijn voor de anderen (May, 2013). Zo zullen acties ter waardering (zoals een projectevaluatie) invloed hebben op de begripsvorming over de meerwaarde en werking van lerend evalueren, en mogelijk leiden tot veranderingen in de toepassing.

**Tabel 2.2 De kernmechanismen van NPT nader toegelicht (May & Finch, 2009; Verwoerd et al., 2020).**

<b>Kernmechanisme</b>	<b>Toelichting</b>
<b>Begripsvorming (sensemaking)</b>	Handelingen en acties ter differentiatie van de innovatie t.o.v. andere reeds bestaande methodieken, en begripsvorming m.b.t. het doel en de meerwaarde
<b>Draagvlak (engagement)</b>	Handelingen en acties om betrokkenheid en draagvlak voor de innovatie aan te moedigen en de innovatie te legitimeren, zoals een (informele) community of practice ontwikkelen en behouden
<b>Toepassing (enactment)</b>	Handelingen en acties t.b.v. het uitvoeren van de methodiek, zoals tools en instrumenten, maar ook de ontwikkeling van gedeelde principes en vaardigheden nodig om de methodiek met kwaliteit te kunnen toepassen. Het toe-eigenen en aanpassen van de methodiek zodat deze past binnen de organisatiecontext behoort hier ook bij
<b>Waardering (appraisal)</b>	Handelingen en acties om de uitvoering en effecten van de methodiek te kunnen waarderen t.b.v. het aanpassen en systematiseren daarvan, zoals formele evaluaties, maar ook meer informele (team)reflecties

Door voor elk mechanisme in kaart te brengen wat voor handelingen er worden ondernomen en waar het eventueel spaak loopt, kan het proces van normaliseren in beeld worden gebracht. Bovendien kan op deze wijze tot een overweging worden gekomen op welke mechanisme aanvullende acties eventueel nodig zijn (May et al., 2011; Murray et al., 2010).

### 3. Onderzoeksaanpak

#### 3.1 PBL als paradigmatische casus

We presenteren het PBL als een zogeheten paradigmatische casus (Flyvbjerg, 2006) van een beleidsonderzoeksorganisatie die bezig is met zijn evaluatierepertoire te innoveren, waarbij meer reflexieve benaderingen worden toegevoegd aan de bestaande praktijk van meer traditionele benaderingen. Door via deze bril naar de organisatie te kijken kunnen niet alleen lessen worden getrokken voor het PBL zelf, maar ook lessen die relevant zijn voor vergelijkbare (beleidsonderzoeks)organisaties.

PBL voert lerende evaluaties uit sinds 2014.<sup>2</sup> Dat jaar kwam vanuit het ministerie van Landbouw, Natuur en Voedselveiligheid (LNV) en het Interprovinciaal Overleg (IPO) het verzoek om het natuurbeleid dat vloeide uit het Natuurpact (2013) tot en met 2027 *ex durante* te evalueren. De toen recente decentralisatie van de ontwikkeling en de uitvoering van het natuurbeleid van het Rijk naar de provincies (gepaard met de nodige spanningen tussen deze overheden), en de verbreding van de ambities van dat natuurbeleid maakte dat er door Rijk en provincies werd gezocht naar een constructieve wijze van evalueren die tijdig inzage zou bieden in de voortgang van het beleid, zodat bijsturen mogelijk werd (zie ook Folkert et al., 2018; PBL & WUR, 2017). Lerend evalueren werd aangedragen als passende methodiek. De lerende evaluatie van het Natuurpact (LEN) wordt uitgevoerd in samenwerking met Wageningen University & Research in driejaarlijkse cycli en loopt nog steeds (met als beoogd eindpunt 2027).

Mede dankzij het succes van de LEN, kwam in 2018 het verzoek om een lerende evaluatie van het Interbestuurlijke Programma Vitaal Platteland (LEVP). Rond dezelfde periode kwam er een ander verzoek van LNV voor een lerend kennisprogramma voor de Regio Deals voor brede welvaart (RD). Lerend onderzoek raakt daarmee steeds meer verankerd in PBL's repertoire, en maakt de organisatie een interessante casus om het proces van normaliseren en professionaliseren nader onder de loep te nemen. Embedded case study aanpak en actieonderzoek

Om de praktijk van lerend evalueren bij het PBL in beeld te brengen is een embedded case study aanpak toegepast (Yin, 2012). We hebben daarbij diverse cases (namelijk lerende evaluaties) binnen eenzelfde context (het PBL) geselecteerd om deze met elkaar te kunnen vergelijken. Daarnaast is de studie niet alleen retrospectief ingestoken, maar ook actiegericht - twee van de drie lerende evaluaties die fungeerden als casus waren nog in gang tijdens de toepassing van deze studie, en voor de derde werden de plannen voor de volgende periode gevormd. Middels actieonderzoek (Reason & Bradbury, 2001) hebben we niet alleen uitdagingen van projectteams met de toepassing van lerend evalueren in beeld gebracht en gepoogd deze beter te begrijpen, maar door dit gezamenlijk te doen hebben we ook ingezet op het ondersteunen van de projectteams met deze uitdagingen het hoofd te bieden.

### 3.2 Casusselectie

Om inzage te krijgen in hoe lerend evalueren plaatsvindt bij het PBL zijn drie lerende evaluatie projecten bij het PBL geselecteerd als cases: de lerende evaluatie van het Natuurpact (de LEN), het lerend onderzoeksprogramma Regio Deals voor brede welvaart (RD) en de lerende evaluatie van het Interbestuurlijke Programma Vitaal Platteland (IBP VP). Tabel 3.1 geeft de kenmerken van deze cases weer, waaronder een beschrijving van het beleidsprogramma waar de lerende evaluatie zich respectievelijk op richtte, wanneer het project plaatsvond en wie de formele verzoekers waren.

---

<sup>2</sup> Overigens kent de roep om en toepassing van meer participatieve en reflexieve onderzoeksmethoden een langere geschiedenis bij het PBL, die bovendien goed is gedocumenteerd. Zie o.a. Hage et al., 2010; Pesch et al., 2012; Petersen et al., 2011.



Een selectiecriteria voor deze studie was dat deze projecten ten tijde van de start van deze studie alle drie formeel geormerkt waren als lerende evaluatie.<sup>3</sup> Een ander selectiecriteria was de betrokkenheid van één of meerdere huidige auteurs bij het project. Dit maakte het mogelijk om als *participant observer* (Schmidt, 2016) de praktijk te aanschouwen en deze zo beter te benaderen in het onderzoek.

**Tabel 3.1** Overzicht van kenmerken van de drie geselecteerde lerende evaluaties die dienden als casus in dit onderzoek.

Casus	1. De lerende evaluatie van het Natuurpact	2. Het lerend onderzoeksprogramma Regio Deals voor brede welvaart	3. De lerende evaluatie van het interbestuurlijk programma Vitaal Platteland
<b>Beleidsprogramma en ambities</b>	Het Natuurpact (2014), gericht op het stoppen van het verlies van biodiversiteit, en het versterken van de relaties tussen natuur, samenleving en economie.	Het Regio Dealprogramma, gericht op het sluiten van Deals tussen Rijk en regio om brede welvaart op regionaal niveau te versterken	Het Interbestuurlijke Programma Vitaal Platteland, gericht op het versterken van interbestuurlijke samenwerking en het realiseren van vitaal platteland
<b>Sturingsarrangement</b>	Gedecentraliseerd natuurbeleid waarbij de provincies verantwoordelijk zijn voor de ontwikkeling en uitvoering van beleid, en het Rijk voor het realiseren van de VHR doelstellingen; in toenemende mate sprake van samenwerking met maatschappelijke partijen en provincies	Interbestuurlijk partnerschap tussen het Rijk (met name LNV en BZK) en 30 Nederlandse regio's. Wie of wat de regio is, verschilt van regio tot regio.	Interbestuurlijk partnerschap tussen Rijk, provincies, gemeenten en waterschappen en 15 gebieden in Nederland.
<b>Tijdspad</b>	2014-heden (bestaat uit driejaarlijkse cycli waarvan de eerste en tweede reeds zijn afgerond)	2018-2021	2019-2021
<b>Formele verzoekers van de lerende evaluatie</b>	2 verzoekers: het Ministerie van Landbouw, Natuur en Voedselkwaliteit (LNV) en de provincies, vertegenwoordigd door het Interprovinciaal Overleg (IPO)	1 verzoeker: LNV	4 verzoekers: LNV, IPO, Vereniging der Nederlandse Gemeenten; (VNG), Unie van Waterschappen (UvW)

<sup>3</sup> Gaandeweg is het Regio Deal onderzoeksprogramma geherformuleerd als lerend onderzoeksprogramma, omdat een te expliciete evaluatie component niet wenselijk werd bevonden door de verzoekers, het ministerie van LNV (er was alsnog wel sprake van een analyse van de werking van beleid). Om de tekst begrijpelijk te houden, refereren we naar alle cases als lerende evaluaties. Onder de resultaten bespreken we in meer detail te uitdagingen met beperkt draagvlak voor lerend evalueren onder betrokken beleidspartijen.

### 3.3 Dataverzameling en -analyse

Het casusonderzoek bestond uit een documentenanalyse van de belangrijkste stukken van de projecten (zoals evaluatiekaders, projectplannen en tussen- en eindrapportages) en in totaal negen semigestructureerde interviews met projectteamleden, projectleiders en supervisors. De interviews waren gericht op het naar boven brengen van de ervaringen met lerend evalueren, wat goed ging en wat minder, en hoe er is gekozen om met bepaalde uitdagingen en knelpunten om te gaan. De interviews zijn met toestemming opgenomen en getranscribeerd ten behoeve van analyse. Daarnaast zijn tijdens onze betrokkenheid aan de cases observaties bijgehouden in field notes. Vervolgens zijn de data gecodeerd op uitdagingen die de projectteams tegenkwamen met het uitvoeren van hun lerende evaluatie. Deze uitdagingen zijn vervolgens gecategoriseerd naar de mechanismen uit het NPT kader (i.e., uitdagingen met begripsvorming, draagvlak, toepassing en waardering).

### 3.4 Reflectiesessies

Op basis van de analyses is eerst voor elk project een casusbeschrijving gemaakt die de basis heeft gevormd voor diverse reflectiesessies waar de belangrijkste kenmerken en uitdagingen van dat project zijn getoetst en besproken (voor de LEN had dit het karakter van een begeleide zelfevaluatie, omdat het project al was afgerond). Afsluitend vond er een bredere en overkoepelende reflectiesessie plaats waarbij (bijna) alle betrokkenen bij de drie projecten gezamenlijk de praktijk van lerend evalueren in beeld hebben gebracht en bediscussieerd. Daarbij nam eveneens de chieft scientist van het PBL deel, in het kader van methodische kwaliteitsborging. Tijdens deze reflectiesessie werd ook gesproken over gedeelde uitdagingen en wat nodig is om de methodiek te normaliseren binnen het PBL en het lerend vermogen te vergroten.

### 3.5 Validatie

Naast de gebruikelijke member checks van de interviews en validatie van onze bevindingen tijdens de diverse reflectiesessies, is ook aanvullende interne en externe kritiek georganiseerd op de praktijk van lerend evalueren bij het PBL. Intern zijn twee gesprekken met collega's met een meer uitgesproken sceptische blik op de methodiek gevoerd om eventuele blinde vlekken te ondervangen. Extern zijn twee partijen benaderd die eveneens lerend evalueren toepassen (de Erasmus Universiteit Rotterdam en het RIVM). Door deze gesprekken was het mogelijk om de PBL praktijk nader te differentiëren en specificeren (deelvraag 1) en om zicht te krijgen op specifieke PBL uitdagingen, en wederom eventuele blinde vlekken te identificeren (deelvraag 2).

## 4. Lerend evalueren: de PBL praktijk

Voordat we ons richten op de uitdagingen die projectteams bij het PBL tegenkomen met het toepassen en inbedden van lerend evalueren in het evaluatierepertoire, beschrijven we eerst hoe de praktijk van lerend evalueren eruitziet bij de organisatie.

## 4.1 Beschrijving van de PBL praktijk

### **Ontwikkeling van 'richtinggevende principes' als handvatten voor lerend evalueren**

Om houvast te krijgen in wat lerend evalueren is en hoe deze methodiek met kwaliteit toe te passen, heeft het PBL in 2015 opdracht gegeven – in het kader van de LEN – aan VU Athena om een literatuurstudie uit te voeren en een theoretisch kader te ontwikkelen voor lerend evalueren. VU Athena heeft voor deze studie geput uit tal van bronnen over participatieve evaluatiemethoden die worden voorgesteld voor de evaluatie van complexe (beleids)programma's, die zich al dan niet specifiek richten op transitieopgaven en systeeminnovatie (Van Veen et al., 2016). Zodoende is het kader dat PBL hanteert voor lerend evalueren o.a. geënt op Reflexieve Monitoring in Actie (van Mierlo et al., 2010), realist evaluation (Pawson & Tilley, 1997), utilization-focused evaluation (Patton, 2000), developmental evaluation (Patton, 1994), responsive evaluation (Stake & Abma, 2005) en learning(-focused) evaluation (Edelenbos & van Buuren, 2005). Al deze methodieken delen dat ze participatief zijn, plaatsvinden parallel aan hetgeen dat wordt geëvalueerd en dat ze zich richten op het vergroten van het lerend vermogen van diegenen wiens werk wordt geëvalueerd ten behoeve van de kwaliteit en de bedoeling van dat werk.

Bij het PBL wordt de functie van lerend evalueren gedefinieerd als het vergroten van het lerend vermogen van beleidsbetrokkenen om de opgave aan te gaan, door tijdens de ontwikkeling en uitvoering van het beleid kennis te coproduceren over de werking en voortgang van (de sturing op) dat beleid. Verder is de methode geconceptualiseerd als bestaande uit een zestal samenhangende 'richtinggevende' principes welke worden doorontwikkeld naarmate de ervaring met lerend evalueren groeit. De principes zijn richtinggevend in de zin dat ze altijd nog operationalisering vergen in een specifieke context. Het is aan het projectteam deze operationalisering te verzorgen:

- i. Lerend evalueren dient om het lerend vermogen in het beleid te vergroten in het gezicht van weerbarstige opgaven, en is een geschikte methodiek voor complexe, interbestuurlijke beleidsprogramma's die doelzoekend van aard zijn en een expliciete leerambitie hebben (al dan niet vergezeld van een geëxpliciteerde transitie-opgave).
- ii. Lerend evalueren vindt plaats tijdens het beleidsproces (*ex durante*) en is iteratief en emergent van aard – het onderzoek is responsief naar de ontwikkelingen in het beleid en wat voor typen inzichten die nodig zijn om knelpunten op te lossen en kansen te benutten. Op basis hiervan worden verschillende tools ingezet om het lerend vermogen te voeden.
- iii. Lerend evalueren betreft maatwerk: het evaluatiecriterium (van doelbereik tot transformerend vermogen, en combinaties daartussen) en de inhoudelijke afbakening worden afgestemd op de behoeften in de praktijk. Daarbij worden evaluatiemethoden gericht op beleidsverantwoording én op beleidsleren gecombineerd. Verantwoorden en leren zijn daarin geen tegenpolen, maar juist elkaar versterkende functies.

- iv. Lerend evalueren is een participatieve beleidsevaluatiemethodiek waarbij relevante beleidsactoren een actieve rol vervullen en eigenaarschap ervaren over het evaluatieonderzoek. Wie relevant is, is situatie- en contextafhankelijk. Diversiteit in betrokken beleidsactoren is daarbij essentieel om zo tot nieuwe inzichten te komen, blinde vlekken te identificeren en creatieve oplossingen te ontdekken voor complexe, meervoudige problemen.
- v. Lerend evalueren vraagt om een passende teamsamenstelling waarin beleidsinhoudelijke en bestuurskundige expertise zijn vertegenwoordigd, om zo tot integrale inzichten te komen. Expertise op vlak van interactieve en participatieve onderzoeksmethoden gericht op leren is een *must-have*.
- vi. Lerend evalueren vraagt om rolzuiverheid en rolvastheid van zowel beleidsonderzoekers als betrokken beleidsactoren. Door transparant en open te zijn over werkzaamheden, keuzes en prioriteringen ten aanzien van de inhoud en uitvoering van de lerende evaluaties, borgen onderzoekers de geloofwaardigheid, onafhankelijkheid en beheersbaarheid van het onderzoek

### ***Lerend evalueren als gefaseerde, participatieve onderzoeksprogramma's***

Voor de lerende evaluaties bij het PBL kunnen verschillende fases worden onderscheiden. Omdat elke lerende evaluatie maatwerk betreft om goed aan te sluiten op het voorliggende beleidsprogramma gegeven de context en de leerbehoeften die daar leven, is er aanvullend steeds sprake van een lange verkennende aanloopfase, langer dan gebruikelijk bij PBL-onderzoek. De resterende fases wijken niet per se ver af van meer gangbaar (evaluatie)onderzoek. Ze worden onderscheiden omdat projectteams per fase bekijken hoe voor die fase de interactie met de beleidspraktijk het beste vorm kan worden gegeven – hoe wordt die fase 'lerend' gemaakt? Bovendien kunnen de fases 1 t/m 5 tijdens een lerende evaluatie meerdere malen worden doorlopen (samen vormen deze fases één leercyclus) om zo de 'opbrengst' van het leren te vergroten. De volgende activiteiten kunnen worden onderscheiden:

- 0. Verkennende scopingfase
- 1. Gezamenlijke ontwikkeling van het evaluatiekader
- 2. Gezamenlijke dataverzameling en -analyse
- 3. Gezamenlijke interpretatie en conclusies trekken
- 4. Gezamenlijke formuleren van handelingsperspectieven
- 5. Schrijven van tussen- of eindrapportage

Het woord 'gezamenlijk' verwijst naar (meer of minder intensieve) samenwerking tussen het projectteam en de beleidsbetrokkenen. Voor deze samenwerking wordt de beleidspraktijk en de verschillende partijen die daarin een rol spelen in beeld gebracht tijdens de verkennende fase, zodat keuzes kunnen worden gemaakt over wie wanneer en hoe te betrekken (wie is de primaire doelgroep? Wie vormen een meer buitenste schil?). Dit betrekken vindt vervolgens bijvoorbeeld plaats via regelmatig overleg met de verzoekende partijen, interviews en focusgroepen, en interactieve workshops waarvoor beleidsbetrokkenen worden uitgenodigd om met de onderzoekers en met elkaar in gesprek te gaan. De evaluatoren blijven overigens

eindverantwoordelijk voor de inhoud en aanpak van de evaluatie, ter borging van onafhankelijkheid.

De lerende evaluaties bij het PBL worden vormgegeven als onderzoeksprogramma's waarvoor deelprojecten vanuit verschillende perspectieven verschillende delen van het beleidsprogramma belichten en in samenhang tot conclusies over dit beleid komen. Er is daarbij telkens sprake van een *mixed methods* aanpak; van empirisch, kwalitatief casusonderzoek, gecombineerd met meer kwantitatief of theorie-gedreven onderzoek.

### **Leren en verantwoorden: een diverse invulling**

Leren ten behoeve van verbetering en verandering staat centraal tijdens een lerende evaluatie. Desalniettemin zijn inzichten over de doeltreffendheid en doelmatigheid van (de sturing op) het beleid een relevante bouwsteen om de werking te begrijpen. Met een lerende evaluatie komen leren en verantwoorden dus samen; de ogenschijnlijke spanning tussen deze twee functies van evalueren wordt opgelost door ze te erkennen als elkaars verlengden. Dit vergt een goede samenwerking, openheid en vertrouwen tussen beleidsbetrokkenen en beleidsonderzoekers.

Hoe vervolgens de evaluatiefuncties leren en verantwoorden in één evaluatieonderzoek worden gecombineerd, verschilt van project tot project. Er wordt daarvoor een afweging gemaakt tussen wat de beleidspraktijk behoeft en wat kansrijke inzichten oplevert. Er is dus sprake van maatwerk. Omdat het combineren van leren en verantwoorden zo karakteristiek is voor de methodiek, staan we in deze paragraaf nader stil bij de vraag hoe deze combinatie gestalte krijgt in de projecten bij het PBL.

In de lerende evaluaties bij het PBL krijgt beleidsleren door beleidsbetrokkenen invulling via het tussentijds delen van inzichten door onderzoekers over de werking en voortgang van (de sturing op) het beleid, en door vervolgens het gezamenlijk gesprek te voeren over de betekenissen van deze inzichten. Ook wordt leren gevoed door gezamenlijk te reflecteren op de gehanteerde beleidstheorie<sup>4</sup> (al dan niet door die eerst te helpen expliciteren). Daarnaast krijgt leren invulling via het faciliteren van het uitwisselen van ervaringen tussen beleidsbetrokkenen, bijvoorbeeld tijdens workshops. Hoewel in wat mindere mate, wordt tijdens lerend evalueren beleidsleren ook gevoed door de vaardigheden van beleidsbetrokkenen te (helpen) ontwikkelen, door workshops te organiseren waar betrokkenen leren zelf een lerende monitoring op te zetten om hun beleid mee te volgen. Tenslotte wordt in enkele lerende evaluaties ook aandacht geschonken aan het analyseren en vergroten van het lerend vermogen van het beleid. Lerend vermogen wordt daarbij door onderzoekers als randvoorwaarde gezien voor lerend en adaptief beleid.

---

<sup>4</sup> Het logisch geformuleerde geheel van veronderstellingen en verwachtingen over de relaties tussen de inzet van middelen en instrumenten en het realiseren van de geformuleerde doelstellingen en ambities.

**Tabel 4.1 'Leren' door beleidsbetrokkenen wordt tijdens lerende evaluaties op verschillende wijzen gevoed**

<b><i>Beleidsbetrokkenen leren tijdens lerend evalueren door middel van...</i></b>
...tussentijds delen van inzichten over de werking en voortgang van (de sturing op) het beleid door onderzoekers en hierover het gesprek voeren
...gezamenlijk reflecteren op de gehanteerde beleidstheorie (al dan niet door die eerst te helpen expliciteren)
...faciliteren van het uitwisselen van ervaringen tussen beleidsbetrokkenen
...de vaardigheden van beleidsbetrokkenen (helpen) ontwikkelen
...analyseren en vergroten van het lerend vermogen van het beleid

Wat betreft verantwoording is de PBL praktijk divers; niet alleen doelbereik maar juist ook andere evaluatiecriteria worden gehanteerd. Zo wordt tijdens de ene lerende evaluatie inzichten ontwikkeld op het vlak van doelbereik, gecombineerd met lessen voor het vergroten van dit bereik, terwijl een andere het transformatie potentieel van het beleidsarrangement als evaluatiecriteria hanteert. Ook is de effectiviteit van een nieuwe interbestuurlijke samenwerking als criterium gebruikt.

Welk evaluatiecriteria wordt gehanteerd om het beleid op te beoordelen (*verantwoorden*) hangt af van de vragen in behoeften die leven in de beleidspraktijk én van welke kennis in het licht hiervan nog ontbreekt. Er wordt dus per lerende evaluatie afgewogen wat de niche is van deze lerende evaluatie en wat het meest kansrijk is om het onderzoek op te richten, in het licht van de voorliggende opgave.

#### **4.2 'Verantwoorden' vindt plaats op diverse aspecten, niet (per sé) doelbereik**

<b><i>Verantwoording vindt plaats middels...</i></b>
...aantonen van voortgang op beleidsdoelen
...aantonen van transformatiepotentieel
...aantonen van effectiviteit van nieuwe interbestuurlijke samenwerking

#### **4.2 Uitdagingen met het toepassen en inbedden van lerend evalueren bij het PBL**

In de drie cases komen we een aantal gedeelde uitdagingen tegen in de uitvoering van lerend evalueren. Deze beschrijven we hieronder, gestructureerd volgens de mechanismen van normaliseren uit het NPT-kader. Het betreffen uitdagingen die zowel door onze gesprekspartners die direct betrokken waren bij deze cases zelf zijn benoemd, als die door ons zijn geïdentificeerd.

##### ***Uitdagingen met betrekking tot begripsvorming***

---

*Begripsvorming* draait om de mate waarin er een gedeeld beeld is onder betrokkenen over wat lerend evalueren is en doet, hoe de methodiek differentieert van andere evaluatiebenaderingen, wat de meerwaarde is en globaal hoe het wordt uitgevoerd.

---

Met betrekking tot het NPT mechanisme *begripsvorming* identificeren we twee uitdagingen. Ten eerste constateren we dat het een uitdaging is voor de projectteams om alle neuzen dezelfde kant op te krijgen en een gedeeld begrip te ontwikkelen voor wat het doel is van de lerende evaluatie en hoe dit zal worden bereikt. Ten tweede constateren we dat er ten behoeve van begripsvorming, beperkte documentatie en materiaal beschikbaar is binnen de organisatie om deze begripsvorming te voeden. We lichten beide toe.

#### *Veronderstellingen en verwachtingen ten aanzien van lerend evalueren lopen uiteen*

We observeren dat er bij aanvang van de lerende evaluaties in eerste instantie sprake lijkt van een gedeeld beeld onder de leden van de verschillende projectteams over wat lerend evalueren 'is' en wat de motivatie is voor een lerende aanpak in het project. Tijdens de uitvoering blijkt dit beeld echter (te) oppervlakkig, wat aanleiding gaf tot verwarring en discussie in het team over de te varen koers.

Onder de projectteamleden leven impliciete en subtiel verschillende veronderstellingen en verwachtingen ten aanzien van lerend evalueren en hoe dit te 'doen', en de rol die zij als onderzoekers vervullen met het onderzoek. De motivatie voor een lerende aanpak verschilt bovendien: voor sommigen draait het om het vergroten van de bruikbaarheid en impact van het onderzoek (een instrumentele blik), terwijl anderen willen bijdragen aan het lerend of transformerend vermogen van het beleid (een transformatieve blik). De tweede is 'radicaler' en voelt voor sommigen ongemakkelijk. In de cases voerde de instrumentele blik overigens de boventoon (met uitzondering van het IBP VP), soms tot onvrede van enkele teamleden. De verschillen in voorkeur voor perspectieven zijn te herleiden naar verschillen in disciplinaire achtergrond en ervaring met participatieve en deliberatieve onderzoeksmethoden.

Hoe is omgegaan met uiteenlopende veronderstellingen en verwachtingen verschilt tussen de projecten, ook naar gelang hoe bewust het betreffende team was van een eventueel tekort aan gedeeld begrip. Er werd (met wisselend succes) ingezet op wederzijdse verdieping om elkaar beter te leren begrijpen. Ook is er in een aantal gevallen voor gekozen om onderzoeksbenaderingen te verdelen over deelprojecten: sommige deelprojecten werden daar door meer participatief en 'lerend' van aard, terwijl andere een meer reguliere benadering volgden.

#### *Beschikbare vastgelegde kennis biedt beperkt handvatten*

Ten behoeven van gedeelde begripsvorming over wat lerend evalueren is en hoe het zou kunnen worden toegepast, is er door de projectteams gezocht naar aanvullende informatie over de methodiek. Binnen PBL is er enige kennis over lerend evalueren vastgelegd naar aanleiding van de ervaringen die zijn opgedaan met de LEN1. Zo is er een theoretisch kader ontwikkeld voor lerend evalueren door VU Athena in opdracht van PBL (tevens ten dele beschreven onder paragraaf 4.1 in deze notitie, naar Van Veen et al., 2016) en zijn procesprincipes opgetekend voor lerend evalueren ten behoeve van de LEN2 (intern document). Verder zijn er diverse rapporten en artikelen gepubliceerd over de uitdagingen (Folkert et al., 2018) en doorwerking van de aanpak (Verwoerd et al., 2017; 2018; 2019). Ook is in 2018 een FAQ over de methodiek opgesteld (PBL, 2018), en er is in 2019 een filmpje gemaakt waarin de

methodiek wordt toegelicht (PBL, 2019). Hoewel deze bronnen helpen om een basisbegrip te ontwikkelen van lerend evalueren, bieden ze beperkt praktische handvatten om echt aan de slag te gaan. Dit kans deels worden verklaard doordat bovengenoemde bronnen vooral gericht zijn op de beleidswereld, en op het vergroten van bekendheid en draagvlak voor de aanpak onder beleidsbetrokkenen (en dus meer worden ingezet ten behoeve van het tweede mechanisme voor normaliseren, *draagvlak*).

### ***Uitdagingen met betrekking tot (het organiseren van) draagvlak***

---

*Draagvlak* definiëren we hier als de mate waarin beleidsonderzoekers (en mogelijk supervisors en managers) en beleidsbetrokkenen lerend evalueren een legitieme methodiek vinden en deze ondersteunen en goedkeuren. Dit bepaalt mede hun bereidwilligheid om actief betrokken te zijn.

---

Met betrekking tot het NPT mechanisme *draagvlak* voor lerend evalueren en het organiseren daarvan, zien we enerzijds een uitdaging wat betreft het draagvlak voor lerend evalueren onder PBL collega's – nodig voor in het algemeen ondersteuning en bijvoorbeeld betrokkenheid van collega's uit de verschillende (inhoudelijke) onderzoeksafdelingen van de organisatie, nodig om een inhoudelijk robuuste evaluatie uit te voeren – en anderzijds met draagvlak onder de beoogde deelnemende beleidsbetrokkenen, zowel wat betreft verzoekers als meer indirect betrokkenen.

#### *Draagvlak onder PBL collega's: lerend evalueren als heel nieuwe tak van sport*

De teams hebben wisselend draagvlak ervaren onder PBL collega's voor lerend evalueren. Tijdens PBL-seminars reageerden collega's soms (zeer) kritisch op de methodiek, met name in relatie tot de onafhankelijkheid en objectiviteit van betrokken onderzoekers (en de geloofwaardigheid van de organisatie in het verlengde daarvan). Ook wordt lerend evalueren als zeer tijdsintensief gezien, en lopen de meningen uiteen of het PBL de juiste partij is om dit soort type onderzoek uit te voeren of dat het beter aan anderen kan worden overgelaten (zoals aan consultants of beleidsambtenaren zelf). Dat sommigen lerend evalueren niet zien als serieuze en gedegen evaluatiemethodiek betekent in de praktijk dat het soms puzzelen is naar hoe een projectteam samen te stellen waarin ook voldoende expertise aanwezig is over de beleidsinhoud van het respectievelijke beleidsprogramma. In de regel zijn dit onderzoekers zonder ervaring met lerend evalueren. Een projectteam met alleen onderzoekers met expertise in lerende en participatieve methoden en bestuurskunde loopt het risico de beleidsinhoud van het beleidsproces los te knippen, met verlies van kwaliteit van het onderzoek in de inzichten als gevolg.

Om met de kritiek om te gaan en om draagvlak te ontwikkelen zetten projectteams in op goede borging van onafhankelijkheid en het uitdragen van hoe dit gebeurt binnen de organisatie. Aanvullend is ingezet op het verankeren van lerend evalueren in wetenschappelijke literatuur, om de wetenschappelijke robuustheid van de methodiek te benadrukken. Daarnaast zetten projectleiders en supervisors zich in als ambassadeurs voor lerend evalueren om bekendheid en draagvlak binnen PBL te ontwikkelen.



Als gevolg van deze activiteiten is het draagvlak voor de methodiek met de jaren toegenomen binnen PBL. Dit blijkt bijvoorbeeld uit de afname van kritische uitlatingen in de wandelgangen en tijdens seminars, en de vanzelfsprekendheid waarmee lerend evalueren wordt gezien als kansrijke methode.

#### *Betrokkenheid van verzoekers: onbekend maakt onbemind?*

Voor een succesvolle lerende evaluatie is draagvlak onder de verzoekers een vereiste, ten behoeve van hun actieve betrokkenheid en bereidheid tot leren. In de cases zien we dat deze zaken niet vanzelfsprekend zijn, zelfs niet wanneer (een deel van) de verzoekende partijen zelf hebben gekozen voor een lerende aanpak en onderschrijven te willen leren gedurende het beleidsproces. Beperkte betrokkenheid leidt tot problemen in de uitvoering van lerend evalueren. Bijvoorbeeld omdat betrokken beleidsactoren geen inzage in hun beleidsplannen willen geven en niet actief deel willen nemen, uit zorg om afrekening.

Om deze uitdaging aan te gaan hebben de teams ingezet op het ontwikkelen van bekendheid over lerend evalueren, en aan onderling vertrouwen tussen onderzoekers en de beleidspraktijk. Bijvoorbeeld in de vorm van bilaterale gesprekken om relaties op te bouwen, en middels het mobiliseren van bestuurders als ambassadeurs om de lerende aanpak te legitimeren (zie ook Folkert et al., 2018). Ook is ingezet op meer structurele overleggen met de verzoekers om zo te werken aan een vertrouwensband en hen meer mee te nemen in de ontwikkelingen van het onderzoek. Ook zijn in de projecten diverse gesprekken gevoerd op 'managerial level' (tussen de PBL supervisor en de relevante counterparts in de beleidsprogramma's) om draagvlak voor de lerende aanpak op bestuurlijk niveau te vergroten.

#### *Betrokkenheid van regionale actoren is lastig zonder duidelijke vertegenwoordiging*

Voor alle cases was sprake van een *interbestuurlijk* beleidsprogramma; de programma's zijn georganiseerd als samenwerkingen tussen Rijk en provincies (LEN), regio's (RD) en gebieden (IBP VP). Voor de lerende evaluaties was het een zoektocht naar hoe de regionale partijen ook te betrekken bij het onderzoek. Dit was uitdagend omdat er geen duidelijk vertegenwoordigend aanspreekpunt was voor de provincies resp. gebieden resp. regio's waarmee afstemming over de inhoud van de lerende evaluatie plaats had kunnen vinden (overigens werd deze voor de LEN wel gevonden in de IPO Werkgroep Natuurbeleid waarin provinciale beleidsambtenaren van de natuurafdelingen vertegenwoordigd zijn). Extra uitdagend was dat de provincies/regio's/gebieden zeer verschillend waren en uiteenlopende evaluatie- en leerbehoeften hadden. Dit maakte het lastig om deze te generaliseren in een beheersbaar onderzoeksontwerp dat goed was aangesloten op de regionale beleidspraktijk.

Wat volgens onze gesprekspartners betrokkenheid genereren onder regionale partijen nader bemoeilijkte, is dat er bij het PBL geen duidelijke strategie voorhanden is voor de rol van de organisatie op regionaal niveau. Dit maakte het lastig om je als projectteam te positioneren als kennispartner richting regionale actoren en goed te kunnen communiceren wat deze partijen redelijkerwijs mogen verwachten van het onderzoek.

## ***Uitdagingen met betrekking tot de toepassing***

---

De *toepassing* van lerend evalueren draait om het uitvoeren van de methodiek, inclusief keuzes voor tools en instrumenten, en de vaardigheden die nodig zijn om de methodiek met kwaliteit te kunnen toepassen.

---

Met betrekking tot het derde mechanisme voor normaliseren volgens NPT, namelijk toepassing oftewel de *uitvoering* van de lerende evaluaties observeren we vier voorname uitdagingen. Deze beslaan meer 'interne' uitdagingen die vooral spelen op team-niveau (1 en 2), en uitdagingen die betrekking hebben op de organisatie (3) en de relatie met de beleidspraktijk (4).

### *Interdisciplinaire kennisintegratie gebeurt niet vanzelf*

Voor elke lerende evaluatie was sprake van een interdisciplinaire aanpak en een interdisciplinaire teamsamenstelling. De gedachte is dat hierdoor een integraal perspectief kan worden geboden op de werking en voortgang van (de sturing op) het beleid, wat als meerwaarde van de methodiek wordt gezien. In elk team was de interdisciplinaire samenwerking uitdagend. Elkaars taal, werkwijzen en kaders leren begrijpen vergt een investering en kent als randvoorwaarde dat onderzoekers open staan voor andere onderzoeksbenaderingen en kwaliteitsstandaarden. Ook vraagt dit om bewustzijn van hun eigen kaders. Interdisciplinair samenwerken is door onze gesprekspartners ervaren als een bij vlagen taai proces waarop initieel niet was geanticipeerd. Het gaf spanning en frustratie. Tegelijk zien we ook dat het energie opwekt wanneer het wél lukt.

Hoewel wordt gedeeld dat interdisciplinaire samenwerking en kennisintegratie nodig is tijdens lerend evalueren, is er hiervoor in de cases beperkt structuur of proces ingericht. Er werd min of meer verondersteld dat het 'vanzelf' gebeurt, aldus onze respondenten. Overigens is er wel geleerd van de ervaringen gedurende de lerende evaluaties. Zo zien we dat er steeds meer aandacht komt voor het interdisciplinaire proces, bijvoorbeeld in de vorm van meer regelmatige sparsessies en gezamenlijke schrijfweken.

### *Gewenste expertise en ervaring nog niet altijd beschikbaar*

Volgens onze gesprekspartners is het in de cases om verschillende redenen een uitdaging geweest om een evenwichtige teamsamenstelling te borgen, met daarin zowel beleidsinhoudelijke en bestuurskundige expertise, als expertise op vlak van participatieve onderzoeksmethoden.

Ten eerste blijkt dat onderzoekers met expertise op het vlak van bestuurskunde en participatieve onderzoeksmethoden niet zomaar voorhanden zijn bij het PBL. Enerzijds zijn dit relatief nieuwe velden voor de organisatie, en is het dus niet vreemd dat de aanstelling van dit type expertises nog aandacht behoeft. Anderzijds wordt het belang van dit type expertise, met participatieve onderzoeksmethoden in het bijzonder, soms onderschat. Het ontwerpen en uitvoeren van een effectieve *lerende* workshop is niet iets wat meer regulier getrainde onderzoekers er 'zomaar' bij kunnen doen. Het is een specifieke expertise, dat specifieke training en ervaring vergt, aldus onze respondenten. Ook vraagt lerend evalueren om uitstekend relatiemanagement;

het managen van de betrokkenheid van verzoekers en andere beleidsactoren (en hun soms uiteenlopende evaluatie- en leerbehoeften) met inachtneming van PBL's autonomie en onafhankelijkheid. Dit vraagt om aanvullende en specifieke vaardigheden. Dat dit type expertise niet altijd op waarde wordt geschat bijvoorbeeld zichtbaar in de tijdelijke aanstellingen die deze onderzoekers veelal hebben. Er zijn hierdoor regelmatig wisselingen waardoor opgedane kennis en ervaring deels verloren gaat. Overigens staat dit op het netvlies van het management en directie, er wordt er reeds ingezet op het bestendigen van tijdelijke aanstellingen naar vaste. Daarnaast wordt er ingezet op meer structurele samenwerking met externe kennispartners, zoals het VU Athena Instituut.

Ten tweede zijn verschillende type expertises bij PBL verdeeld over de afdelingen cq. sectoren. Er is behoefte aan 'sectoroverstijgende samenwerking' zodat onderzoekers met de best aansluitende expertise kunnen worden betrokken op een project, maar de respondenten geven aan dat dit niet vanzelfsprekend lukt. Sector specifieke taken krijgen soms voorrang en experts zijn niet altijd vrij te spelen voor sectoroverstijgende betrokkenheid. Anderzijds wordt aangegeven dat onderzoekers zonder ervaring met lerend evalueren niet zomaar te enthousiasmeren zijn voor dit type onderzoek, vermoedelijk omdat ze niet goed weten wat het inhoudt of hun bedenkingen hebben over de gedegenheid van de methode (zie ook de uitdaging rondom draagvlak onder PBL-collega's).

#### *Wijze van werkprogrammering bij PBL is minder toegesneden op iteratieve en emergente projecten*

Bij het PBL wordt een jaarlijks werkprogramma ontwikkeld waarin voor het volgende jaar de projecten en plannen worden vastgelegd, inclusief capaciteit en budgetten. Hierdoor is het mogelijk overzicht te genereren in de plannen, wie waaraan zal mee zal werken en voor hoeveel uur, en waar mogelijk bruggen kunnen worden geslagen. Dat leent zich niet altijd voor lerende evaluaties die meer iteratief en emergent van aard zijn, en waarvoor er ruimte nodig is om in te kunnen spelen op nieuwe ontwikkelingen in de beleidspraktijk (door nieuwe deelprojecten of producten in werking te stellen, met de bijbehorende capaciteit van medewerkers). Om adaptief en flexibel te kunnen zijn is het daarvoor juist nodig dat niet alles wordt vastgelegd. In het licht hiervan wordt gekeken naar hoe meer flexibiliteit in te bouwen, bijvoorbeeld door plannen meer op hoofdlijnen uit te zetten.

Overigens komt deze behoefte aan perspectief voor hoe de lerende evaluatie er het komende jaar (jaren) uit zal gaan zien ook vanuit de beleidspraktijk, omdat betrokkenen eveneens inzicht willen in wat er van hun wordt verwacht en wat zij mogen verwachten van het onderzoek.

#### *De onderzoeks- en beleidspraktijk kennen verschillende snelheden*

Een andere uitdaging met de toepassing betreft de verschillende snelheden en tijdspaden tussen de onderzoeks- en beleidspraktijk. Tijdens lerende evaluaties wordt veelal casuonderzoek uitgevoerd, een kwalitatieve onderzoeksmethode om verdiepende inzichten te genereren over de praktijk. Het is een capaciteits- en

tijdsintensieve methode, en de respondenten geven aan dat het de uitdaging is om op een (kosten)effectieve en tijdige wijze bruikbare inzichten uit de data te halen.

Wat meespeelt is dat het voor onderzoekers – met name zij die beperkt ervaring hebben met interactief onderzoek – als kwetsbaar wordt ervaren om inzichten al tussentijds te delen. Zij ervaren de kwaliteit van wat ze dan leveren als ondermaats. Het is voor hen prettiger om het onderzoek volledig af te ronden en robuust te maken, voordat ze een inzichtje in hun keuken geven. Beleidsbetrokkenen houden hier een andere kijk op na, en hebben juist behoefte aan vroege inzichten zodat ze kunnen anticiperen en deze tijdig meenemen in het beleidsproces. Hoe onderzoekers en beleidsactoren de kwaliteit van onderzoek beoordelen en waarderen ligt dus uit elkaar.

In de cases leiden deze zaken er toe dat de teams zoeken naar hoe ze sneller tot relevante inzichten kunnen komen, bijvoorbeeld door minder diepgaand onderzoek te doen, of door regelmatig tussentijds inzicht te geven in de voortgang van het onderzoek. We observeren daarbij dat verschillende onderzoekers zich hier steeds comfortabeler mee zijn gaan voelen, en makkelijker het gesprek aan gaan met beleidsactoren over de – mogelijke – betekenissen van hun werk.

### ***Uitdagingen met betrekking tot waardering***

---

Waardering wordt hier in de brede zin bedoeld en omvat alle handelingen en acties om de werking en de impact van lerend evalueren in beeld te brengen, zoals de formele monitoring en evaluatie van de methodiek, maar ook meer informele gesprekken waarbij onderzoekers reflecteren op hun werk. Deze activiteiten voeden de andere drie mechanismen – begripvorming, draagvlak en betrokkenheid – en zijn daarom essentieel voor het normaliseren van lerend evalueren.

---

Wat betreft het vierde mechanisme voor normaliseren volgens NPT, *waardering*, uitent onze gesprekspartners met name de wens om meer 'lerend vermogen', dat wil zeggen, beter kunnen leren van de ervaringen die zijn opgedaan in de cases ten behoeve van toekomstige lerende evaluaties. De dagelijkse gang van zaken maakt volgens hun dat er niet voldoende tijd en ruimte is om te reflecteren op hoe het gaat of om uit te wisselen met andere projectteams.

Voor alle drie de cases was er sprake van een informele vorm van waardering van hun gekozen methodiek. Tussen de bedrijven door vonden gesprekken plaats tussen onderzoekers onderling over de methodiek, wat deze lastig maakt en waar ze over twijfelden, of wat ze juist krachtig vonden. We zien dat er middels dit soort gesprekken meer begrip ontstaat voor lerend evalueren en de toepassing, en voor de impact van onderzoek op beleidsactoren. Dit vergrootte begrip het enthousiasme voor een lerende (evaluatie)aanpak doen toenemen.

Aanvullend vond er in alle drie de cases achteraf een projectevaluatie plaats. Voor de LEN is de evaluatie het meest formeel geweest: zowel in de vorm van begeleide zelfevaluaties, als externe projectevaluaties. VU Athena is daarvoor zowel na de eerste als de tweede periode gevraagd om evaluatieonderzoek uit te voeren naar de kwaliteit en impact van hoe lerend evalueren is toegepast in dat project en hoe dit eventueel beter kan (voor de rapportage hierover zie Brouwers et al., 2021; Verwoerd et al., 2019, 2020). Dit draagt in belangrijke mate bij aan het leren over

de methode en het verbeteren van de kwaliteit daarvan. Opvallend, voor de LEN was ook sprake van monitoring van de toepassing van de methode *tijdens* de uitvoering. Hiervoor werd een logboek bijgehouden om bepaalde beslissingen ten aanzien van de kwaliteit van de methodiek in te documenteren. Ook deze wijze van 'waarderen' van het proces kwam het project ten goede.<sup>5</sup> Tabel 4.1 vat de uitdagingen samen.

**Tabel 4.1. De uitdagingen samengevat per NPT mechanisme.**

Begripsvorming	Draagvlak	Toepassing	Waardering
Uiteenlopende beelden binnen het projectteam over wat lerend evalueren is en doet	Beperkt draagvlak onder PBL-collega's	Interdisciplinaire kennisintegratie is stroperig proces	Waardering vindt vooral informeel plaats
Beperkt beschikbare informatie over de methode	Betrokkenheid en commitment van verzoekers niet vanzelfsprekend	Passende expertise en ervaring niet zomaar beschikbaar	Lessen landen lastig in vervolgprojecten
	Onduidelijk doel en strategie voor betrekken van regionale partijen	Sectoroverstijgend samenwerken niet vanzelfsprekend	Formele waardering vindt plaats na afloop van projecten, minder tijdens het project
		Iteratief en emergente werkwijze minder toegesneden op werkprogrammering	
		Verschillende snelheden van onderzoeks- en beleidspraktijk	

## 5. Concluderende discussie en lessen

### 5.1 Een ontwikkelende praktijk

In deze sectie reflecteren we op de bevindingen aan de hand van literatuur en beantwoorden we de hoofdvraag van deze studie: *Wat zijn lessen voor het normaliseren en professionaliseren van lerend evalueren bij beleidsonderzoeksorganisaties, zoals het PBL?*

Onze bevindingen laten zien dat een lerende evaluatie praktijk – Nb. een geheel van samenhangende en deels geroutineerde activiteiten, zienswijzen, tools en instrumenten (Reckwitz, 2002) – zich aan het ontwikkelen is bij het PBL. Goed beschouwd laten de ontwikkelende praktijk en de daarmee gepaarde uitdagingen een veranderproces zien dat gangbaar is voor innovatieprocessen binnen organisaties. Zo doen bijvoorbeeld Preskill en Torres (2002) uit de doeken dat processen van

<sup>5</sup> Tijdens de afronding van deze studie werd bekend om voor de derde ronde van de LEN en toekomstige lerende evaluaties deze vorm van 'interne monitoring' te formaliseren in de vorm van maandelijkse reflecties met projectleiders op kwaliteit en impact om zo de uitvoering van een lerende evaluatie verder te ondersteunen.

organisatieverandering en -leren verlopen via verschillende fases, namelijk de status quo (i), bewustwording van een tekortkoming van deze status quo (ii), verkenning van alternatieven door een kleine groep pioniers (iii), verspreiding van alternatieve aanpak naar grotere community binnen de organisatie (iv) en tenslotte bestendiging van deze aanpak in de organisatie (v). De praktijk van lerend evalueren bij het PBL bevindt zich naar onze inschatting in fase (iv). Overigens relevant om te benadrukken dat de organisatieverandering die we hier centraal stellen het includeren van lerend evalueren in een breder repertoire van evaluatiebenaderingen betreft, en niet het vervangen daarvan.

Niettemin laten de uitdagingen in paragraaf 4.2 zien dat er verbeterruimte is op elk van de vier NPT mechanismen om het proces van normaliseren en professionaliseren zoals voorgesteld door o.a. May et al. (2018) te versterken. De uitdagingen kunnen worden deels worden verklaard doordat lerend evalueren als evaluatiebenadering ver aflight van wat door velen als 'normaal' wordt gezien voor evaluatieonderzoek. Niet alleen maakt dit dat niet iedereen zomaar 'on board' is bij deze methodiek – zoals collega-beleidsonderzoekers en beleidsbetrokkenen, of de bredere evaluatie community –, het betekent ook dat het systeem er niet op is ingericht. De wijze waarop evaluatie momenteel met name wordt ingezet ter beleidsverantwoording en de 'gouden standaarden' die hiermee gepaard gaan zijn dusdanig geïnstitutionaliseerd dat er weinig ruimte is om hier vanaf te wijken, een fenomeen dat niet alleen in Nederland speelt (Van der Steen et al., 2018), maar ook andere landen in het Westen (o.a. Chouinard, 2013; Nieminen & Hyytinen, 2015). Aanvullend maakt de nieuwigheid van de methodiek dat het nodige expertise nog in ontwikkeling is: de beleidsonderzoekers bij het PBL bevinden zich nog volop in hun eigen leerproces. Dit wordt als kenmerkend gezien voor de totstandkoming van meer interactieve onderzoekspraktijken, en verschillende studies onderschrijven het belang van (het ondersteunen van) leren-door-doen door (beleids)onderzoekers en (beleids)betrokkenen om met elkaar kennis te coproduceren (zie o.a. Pohl et al., 2010; Roux et al., 2010). Dat verschillen in hoe kwaliteit wordt beoordeeld en wat urgent wordt gevonden, en in tijdspaden en taal tussen beleidsonderzoekers en beleidsbetrokkenen zo'n samenwerking bemoeilijkt en vraagt om gezamenlijk leren (Choi et al., 2005), bevestigen de ervaren uitdagingen van de projectteams uit de cases.

## 5.2 Lessen voor normaliseren en professionaliseren

Tenslotte komen we tot lessen ten behoeve van het normaliseren en professionaliseren van lerend evalueren bij PBL. Omdat de vier mechanismen van NPT dusdanig in samenhang functioneren maken we de bewuste keuze om de conclusies en lessen niet ook te structureren volgens deze mechanismen, maar om juist integrale lessen te formuleren die betrekking hebben op meerdere mechanismen gelijktijdig. Hiervoor putten we inspiratie uit het veld van organisatie- en managementwetenschappen en identificeren vier verschillende, interacterende en geneste 'settings' die samen invloed uitoefenen op de praktijk van lerend evalueren, namelijk het individu, het team, de *community of practice* en de organisatie in zijn geheel (Berg Johansen, 2017; Preskill & Torres, 2000). Op basis van de bevindingen werken we drie thema's met bijbehorende lessen uit die zich richten op deze settings (het niveau van het individu laten we daarbij buiten beschouwing).

De lessen die we trekken dienen eveneens ter inspiratie voor andere organisaties.

## ***I. Individuele projectteams: draag zorg voor het interne proces, en kwaliteits- en omgevingsbewustzijn***

We constateren dat individuele projectteams bij het PBL nog bewuster en meer intentioneel kunnen investeren in en sturen op de kwaliteit en impact van hun lerende evaluatie. Hiervoor is zorg nodig voor het interne proces en de samenwerking in het team en aandacht voor kwaliteitsbewustzijn. Aanvullend biedt versterkt omgevingsbewustzijn mogelijkheid om beleidsbetrokkenen goed te betrekken bij de lerende evaluatie (zie ook Kunseler & Verwoerd, 2019).

- *Het interne proces.* Voorafgaand aan een lerende evaluatie is een gesprek nodig over de beelden die er spelen over wat lerend evalueren is en doet, wat de motivatie is voor een lerende evaluatieaanpak gegeven het voorliggende verzoek, en hoe wordt gedacht over PBL's rol. Consensus of unanimitieit is hierbij niet het streven; het expliciteren van perspectieven en het kiezen van een koers wél (zoals ook elders geconstateerd voor inter- en transdisciplinaire kennisprocessen, (Regeer & Bunders, 2009). In aanvulling hierop is het wenselijk om al aan de voorkant van het project een plan voor interdisciplinaire samenwerking te maken. Kennisintegratie gaat niet vanzelf, maar vergt structurele tijdsinvestering en een ingericht proces (Bammer et al., 2020).
- *Kwaliteitsbewustzijn.* Het kwaliteitsbewustzijn kan worden vergroot door meer actief de principes voor lerend evalueren (zoals gepresenteerd in 4.1) te hanteren en hier regelmatig op te reflecteren. Hoewel ze beschikbaar zijn, is het niet altijd standaard routine om de principes te gebruiken om bepaalde keuzes ten aanzien van het onderzoek te maken. Een vorm van structurele en begeleide reflectie kan dit kwaliteitsbewustzijn ondersteunen (zoals ook aangedragen door o.a. Roux et al., 2010). Bovendien kunnen zo gemaakte keuzes worden gedocumenteerd ter naslag. Aanvullend is projectevaluatie met in het specifiek aandacht voor de toepassing van de methodiek wenselijk. De lessen die tijdens zo'n reflectie worden getrokken naderhand breed in de organisatie beschikbaar te maken, wordt tenslotte het lerend vermogen gevoed.
- *Omgevingsbewustzijn.* Vanuit individuele projecten kan meer worden gestuurd op impact door het omgevingsbewustzijn te versterken. Ook hier is regelmatige reflectie noodzakelijk. We achten het daarbij behulpzaam als wordt ingezet op de ontwikkeling van een impactstrategie van de lerende evaluatie waarin de rol en aanpak van het PBL in relatie tot de verschillende betrokken partijen duidelijk wordt gemaakt. Daarin wordt geëxpliciteerd wie de doelgroepen zijn van de lerende evaluatie, hoe wordt verwacht dat zij zullen leren, en op welke wijze beleid eventueel wordt verantwoord en hoe dat invloed zal hebben op de (betrokkenheid van) de doelgroepen. Gezien het responsieve karakter van een lerende evaluatie staan de doelgroepen niet per se vast, maar kunnen zich gaandeweg de lerende evaluatie ontwikkelen (Reed et al., 2014; Reed et al., 2009). Om betrokkenheid verder te borgen is het aanvullend van belang om de rol van de organisatie en het doel van de lerende evaluatie tijdens interactiemomenten structureel toe te lichten, en te zorgen dat beleidsbetrokkenen dit makkelijk kunnen achterhalen (bijvoorbeeld via een website).

## **II. Community of Practice: investeer in lerend vermogen rondom lerend evalueren**

Er is binnen PBL een, hoewel kleine, informele *community of practice* (CoP) aan het ontstaan rondom lerend evalueren: een groep betrokkenen met een gedeelde interesse in de methodiek en verschillende niveaus van ervaring, die elkaar regelmatig treft en daardoor hun expertise en kennis verdiepen (Galarneau, 2002). De *community* is informeel in de zin dat ze elkaar niet structureel treffen om bijvoorbeeld lessen uit te wisselen uit de verschillende lerende evaluaties, of dat ze op andere wijze werken aan een gedeelde agenda. Dit soort type uitwisseling komt onder andere in het geding door de hoge werkdruk en beperkte capaciteit. We zien verschillende kansen voor de community om het lerend vermogen rondom lerend evalueren te versterken.

- *Verstevig de Community of Practice.* CoP's zijn geïdentificeerd als een effectieve wijze van kennismanagement en leerprocessen binnen organisaties (Galarneau, 2002). Door met enige regelmaat te overleggen waarbij het uitwisselen van lessen, bijvoorbeeld van methodologische aard maar ook omtrent bepaalde uitdagingen, centraal staat kan zo'n CoP meer worden verstevigd. Er wordt daarbij aanbevolen om te putten uit de intrinsieke motivatie van praktijkdeskundigen om uit te wisselen, zodat een CoP grotendeels wat organisch ontstaat. Wanneer het te formeel en top down is, is de kans groter dat mensen maar kort lid blijven. Deskundigen dragen daarbij aan dat het belangrijk is dat de inhoud van deze overleggen verder wordt bepaald door wat actueel is en wat zich aandraagt in de praktijk, zodat het overleg relevant is voor hen die daarbij aansluiten (Wenger et al., 2002). Zo'n CoP kan ook de invulling krijgen van intervisie. Wie onderdeel uitmaakt van de CoP is daarbij dynamisch. We achten het kansrijk om bijvoorbeeld ook beleidsonderzoekers uit te nodigen met ervaring met andere vormen van participatief onderzoek dan evaluatie specifiek, en mogelijk ervaringsdeskundigen van andere organisaties.
- *Werk aan de beschikbaarheid van kennis en informatie.* Een toegankelijke kennisbasis met daarin informatie over de achtergrond en theorie van lerend evalueren en die wordt aangevuld naarmate er meer lessen worden opgedaan (bijvoorbeeld middels individuele projectevaluaties) kan een waardevolle bron voor lerend vermogen zijn. Hieronder past bijvoorbeeld ook een handreiking met daarin het proces van lerend evalueren beschreven en de denkstappen die van belang zijn voor een het ontwerpen en succesvol uitvoeren van een lerende evaluatie. Belangrijk is daarbij dat een handreiking geen leidraad of blauwdruk kan zijn; ruimte voor iteratief werken en maatwerk is noodzakelijk (Regeer & Bunders, 2009). Verder is het nodig dat deze kennis en informatie vindbaar is, en beschikbaar en toegankelijk.
- *Maak keuzes en stel randvoorwaarden.* Het toenemend aantal verzoeken voor lerende evaluaties en de beperkte capaciteit bij het PBL vraagt om keuzes te maken en om prioritering. Niet alles kan. Een overweging kan daarbij zijn om, gegeven het capaciteitsprobleem, als PBL nog wel de evaluatiekaders te ontwikkelen maar de uitvoering van lerende evaluaties over te laten aan andere bureaus. Aanvullend laat deze studie een diversiteit aan invullingen van lerend



evalueren zien, gericht op verschillende type vraagstukken. Een overweging is op zijn plaats of het effectiever zou zijn om uitsluitend te richten op een bepaald type vraagstuk en daaraan randvoorwaarden te stellen die de impact van een lerende evaluatie vergroten. Bijvoorbeeld door alleen verzoeken voor lerende evaluaties van transitiebeleid te honoreren, waarvoor een expliciete leerambitie is uitgesproken.

### **III. Organisatieniveau: wees responsief naar de toepassing en doorontwikkeling van lerend evalueren**

Hoewel lerend evalueren steeds meer verankerd raakt in de organisatie, observeren we dat verschillende factoren die op organisatieniveau spelen de succesvolle toepassing van lerende evaluaties bemoeilijken. We identificeren hiervoor een aantal kansen.

- *Organisatiestructuren.* Op enkele vlakken sluiten organisatiestructuren niet goed om de praktijk van lerend evalueren goed in te kunnen bedden. Zo past de wijze van werkprogrammering bijvoorbeeld niet goed bij het iteratieve en emergente karakter van lerend evalueren, en is sectoroverstijgend samenwerking geen vanzelfsprekendheid. Een meer flexibelere werkprogrammering en meer incentive voor samenwerking tussen sectoren (op beide vlakken zijn bij PBL reeds stappen gezet) zou de kwaliteit van lerend evalueren bevorderen omdat dan het iteratieve proces meer ruimte krijgt, en interdisciplinaire samenwerking wordt vergemakkelijkt.
- *Borging van kennisbasis.* De poule van beleidsonderzoekers die PBL in huis heeft voor het uitvoeren van lerende evaluaties is klein. Expertise met transdisciplinaire, deliberatieve en lerende onderzoeksmethodieken beperkt beschikbaar binnen de organisatie, en ervaring die hiermee wordt opgedaan dreigt soms verloren te gaan vanwege tijdelijke aanstellingen. Door tijdelijke aanstellingen te bestendigen in vaste wordt de opgedane ervaring behouden (ook hier zijn relevante ontwikkelingen gaande). Aanvullend zou de kennisbasis kunnen worden versterkt door slim samen te werken met externe partijen. Daarnaast kan er worden ingezet op trainingen en workshops voor participatieve en deliberatieve onderzoeksmethoden voor collega's om de poule zo te vergroten. Omdat lerend evalueren van onderzoekers vraagt om zowel het interne als het externe proces goed te borgen, terwijl vaak complexe omgevingsdynamieken spelen, zijn trainingen voor complex projectmanagement eveneens kansrijk.

Het onderzoek waar deze publicatie op is gebaseerd keek vooral van *binnen naar buiten*: de uitdagingen met lerend evalueren binnen de organisatie, en de invloed van de omgeving daarop. Niettemin onderschrijft dit onderzoek de complexe wisselwerking tussen de evaluatiepraktijk van een beleidsonderzoeksorganisatie, en die van zijn omgeving. De toenemende vraag naar evaluatieonderzoek dat zich richt op het leren binnen beleidsprogramma's sterkt het normaliseerproces. Andersom dragen succesverhalen van lerende evaluaties bij aan deze vraag (en dus het normaliseren van de methode binnen de beleidspraktijk). Dit is belangrijk, omdat het succes van een lerende evaluatie afhangt van de mate waarin deze evaluatiebenadering voet aan de grond heeft in de beleidspraktijk en onder de betrokken deelnemers. Als de

bereidheid tot leren bij hen beperkt is en/of als de structuren die nodig zijn om dit leren te doen plaatsvinden afwezig zijn, wordt de kwaliteit en impact van lerend evalueren serieus beperkt. Enkel met een lerende houding en structuren die dit faciliteren wordt het mogelijk om lerend evalueren uit te voeren op een wijze die het impactpotentieel volledig benut.

Diverse studies laten zien dat hoewel leren in zwang lijkt, dit bij de overheid nog niet altijd navolging krijgt noch tot de verwachte successen leidt (Schulz et al., 2020; van der Steen et al., 2018; Van Der Steen et al., 2014). Leren en lerend evalueren is niet gewoon. Naast een lineaire resultaatgerichtheid en geïnstitutionaliseerde verantwoordingsmechanismen hangt dit in ieder geval ten dele samen met wat geldt als de norm voor beleidsevaluatie bij de evaluatieafdelingen van o.a. ministeries. Zolang in de evaluatie community alleen doeltreffendheid, doelmatigheid en dus het één-op-één aantonen van causale relaties als dé na te streven gouden standaarden voor evaluatie worden gehanteerd, zal het een uitdaging blijven om met lerend evalueren serieus te worden genomen. Lerend evalueren wordt dan gezien als een vrijblijvende, 'laagwaardige' evaluatiemethode, in plaats van als een actiegerichte en impactvolle methodiek die vooruitgang weet te boeken in het gezicht van weerbarstige opgaven. Niettemin, wekken de diverse experimenten met lerend evalueren die plaatsvinden binnen de Rijksoverheid de suggestie dat lerend evalueren een voet tussen deur heeft, en schetsen een gunstig beeld wat dat betreft.

We eindigen deze publicatie dan ook met het onderstrepen van het belang dat lerend evalueren niet alleen normaliseert en professionaliseert bij beleidsonderzoeksorganisaties, maar ook bij de overheden wiens beleid wordt onderzocht en de aanverwante evaluatie community. Ook dáár dient lerend evalueren te normaliseren en professionaliseren als serieuze evaluatiemethodiek voor complexe (transitie-)vraagstukken.

## REFERENTIES

- Bammer, G., O'Rourke, M., O'Connell, D., Neuhauser, L., Midgley, G., Klein, J. T., Grigg, N. J., Gadlin, H., Elsum, I. R., Bursztyn, M., Fulton, E. A., Pohl, C., Smithson, M., Vilsmaier, U., Bergmann, M., Jaeger, J., Merckx, F., Vienni Baptista, B., Burgman, M. A., ... Richardson, G. P. (2020). Expertise in research integration and implementation for tackling complex problems: when is it needed, where can it be found and how can it be strengthened? *Palgrave Communications*, 6(1), 1–16. <https://doi.org/10.1057/s41599-019-0380-0>
- Berg Johansen, C. (2017). What are Institutional Logics - and where is the perspective taking us? In G. Krücken, C. Mazza, R. E. Meyer, & P. Walgenbach (Eds.), *New Themes in Institutional Analysis - Topics and Issues from European Research*. Edward Elgar Publishing.
- Bloemert, H. (2022). *Mapping evaluators and policy actors' beliefs and understandings on impact and quality in reflexive evaluations*. Athena Instituut, VU University Amsterdam, Amsterdam.
- Brouwers, H. J. H., Verwoerd, L., & Klaassen, P. (2021). *Op weg met verantwoord en waardevol leren - Impactstudie van de Lerende Evaluatie Natuurpact 2020*. Athena Instituut, VU University Amsterdam, Amsterdam.
- Choi, B. C. K., Pang, T., Lin, V., Puska, P., Sherman, G., Goddard, M., Ackland, M. J., Sainsbury, P., Stachenko, S., Morrison, H., & Clotey, C. (2005). Can scientists and policy makers work together? In *Journal of Epidemiology and Community Health* (Vol. 59, Issue 8, pp. 632–637). <https://doi.org/10.1136/jech.2004.031765>
- Chouinard, J. A. (2013). The Case for Participatory Evaluation in an Era of Accountability. *American Journal of Evaluation*, 34(2), 237–253. <https://doi.org/10.1177/1098214013478142>
- Edelenbos, J., & van Buuren, A. (2005). The learning evaluation: a theoretical and empirical exploration. *Evaluation Review*, 29(6), 591–612. <https://doi.org/10.1177/0193841X05276126>
- Etienne Wenger, B., McDermott, R., & Snyder, W. M. (2002). Seven Principles for Cultivating Communities of Practice. *HBS Working Knowledge*.
- Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, 12(2), 219–245. <https://doi.org/10.1177/1077800405284363>
- Folkert, R., Verwoerd, L., & Verwest, F. (2018). Lerend evalueren: navigeren tussen verantwoord en leren. *Beleidsonderzoek Online*, 7. <https://doi.org/10.5553/bo/221335502018000007001>
- Funtowicz, S. O., & Ravetz, J. R. (1993). SCIENCE FOR THE POST-NORMAL AGE. *Futures*, 25, 739–755.
- Galarneau, J. (2002). Cultivating Communities of Practice: A Guide to Managing Knowledge. In *Project Management Journal*. <https://doi.org/10.1177/875697280203300312>
- Hage, M., Leroy, P., & Petersen, A. C. (2010). Stakeholder participation in environmental knowledge production. *Futures*, 42(3), 254–264. <https://doi.org/10.1016/j.futures.2009.11.011>

- Houppermans, M. (2018). Operatie Inzicht in Kwaliteit: dwing de cultuuromslag af met een vernieuwde RPE. *Beleidsonderzoek Online*, 0(2). <https://doi.org/10.5553/bo/221335502018000002001>
- Kunseler, E.-M. (2017). *Government expert organisations in-between logics*. VU University Amsterdam, Amsterdam.
- Kunseler, E.-M., & Vasileiadou, E. (2016). Practising environmental policy evaluation under co-existing evaluation imaginaries. *Evaluation*, 22(4), 451–469. <https://doi.org/10.1177/1356389016668099>
- Kunseler, E.-M., & Verwoerd, L. (2019). *Kennis met impact: Reflexief werken bij publieke kennisorganisaties*. PBL, Den Haag.
- Kunseler, E.-M., Verwoerd, L., & Verwest, F. (2020). Inzet op omgevingsbewust en kwaliteitsbewust werken in beleidsonderzoek. *Beleidsonderzoek Online*, november. <https://doi.org/10.5553/bo/221335502020000009001>
- May, C. R. (2013). Towards a general theory of implementation. *Implementation Science*, 8(1), 18. <https://doi.org/10.1186/1748-5908-8-18>
- May, C. R., Cummings, A., Girling, M., Bracher, M., Mair, F. S., May, C. M., Murray, E., Myall, M., Rapley, T., & Finch, T. (2018). Using Normalization Process Theory in feasibility studies and process evaluations of complex healthcare interventions: a systematic review. *Implementation Science*, 13(80), 1–27. <https://doi.org/10.1186/s13012-018-0758-1>
- May, C. R., Finch, T., Ballini, L., MacFarlane, A., Mair, F., Murray, E., Treweek, S., & Rapley, T. (2011). Evaluating complex interventions and health technologies using normalization process theory: development of a simplified approach and web-enabled toolkit. *BMC Health Services Research*, 11(1), 245. <https://doi.org/10.1186/1472-6963-11-245>
- Murray, E., Treweek, S., Pope, C., MacFarlane, A., Ballini, L., Dowrick, C., Finch, T., Kennedy, A., Mair, F., Nio Ong, B., Rapley, T., Rogers, A., & May, C. R. (2010). *Normalisation process theory: a framework for developing, evaluating and implementing complex interventions*. <https://doi.org/10.1186/1741-7015-8-63>
- Nieminen, M., & Hyytinen, K. (2015). Future-oriented impact assessment: Supporting strategic decision-making in complex socio-technical environments. *Evaluation*, 21(4), 448–461. <https://doi.org/10.1177/1356389015606540>
- Owens, S., Rayner, T., & Bina, O. (2004). New agendas for appraisal: Reflections on theory, practice, and research. In *Environment and Planning A* (Vol. 36, Issue 11, pp. 1943–1959). <https://doi.org/10.1068/a36281>
- Patton, M. Q. (1994). Developmental Evaluation. *American Journal of Evaluation*, 15(3), 311–319. <https://doi.org/10.1177/109821409401500312>
- Patton, Michael Quinn. (2000). Utilization-Focused Evaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation Models* (pp. 425–438). Kluwer Academic Publishers.
- Pawson, R., & Tilley, N. (1997). An introduction to scientific realist evaluation. In *Evaluation for the 21st century: A handbook* (pp. 405–418). Thousand Oaks, CA, US: Sage Publications.
- Pesch, U., Huitema, D., & Hisschemöller, M. (2012). A boundary organization and its changing environment: the Netherlands Environmental Assessment Agency, the

- MNP. *Environment and Planning C: Government and Policy*, 30, 487–503. <https://doi.org/10.1068/c10150j>
- Petersen, A. C., Cath, A., Hage, M., Kunseler, E.-M., & van der Sluijs, J. P. (2011). Post-Normal Science in Practice at the Netherlands Environmental Assessment Agency. *Science, Technology & Human Values*, 36(3), 362–388. <https://doi.org/10.1177/0162243910385797>
- Pohl, C., Rist, S., Zimmermann, A., Fry, P., Gurung, G. S., Schneider, F., Speranza, C. I., Kiteme, B., Boillat, S., Serrano, E., Hadorn, G. H., & Wiesmann, U. (2010). Researchers' roles in knowledge co-production: experience from sustainability research in Kenya, Switzerland, Bolivia and Nepal. *Science and Public Policy*, 37(4), 267–281. <https://doi.org/10.3152/030234210X496628>
- Preskill, H., & Torres, R. (2000). The Learning Dimension of Evaluation Use. In *The Expanding Scope of Evaluation Use* (pp. 25–37). <https://doi.org/10.1002/ev.1189>
- Reckwitz, A. (2002). Toward a Theory of Social Practices A Development in Culturalist Theorizing. *European Journal of Social Theory*, 5(2), 243–263.
- Reed, M. S., Stringer, L. C., Fazey, I., Evely, A. C., & Kruijssen, J. H. J. (2014). Five principles for the practice of knowledge exchange in environmental management. *Journal of Environmental Management*, 146, 337–345. <https://doi.org/10.1016/j.jenvman.2014.07.021>
- Reed, Mark S., Graves, A., Dandy, N., Posthumus, H., Hubacek, K., Morris, J., Prell, C., Quinn, C. H., & Stringer, L. C. (2009). Who's in and why? A typology of stakeholder analysis methods for natural resource management. *Journal of Environmental Management*, 90(5), 1933–1949. <https://doi.org/10.1016/j.jenvman.2009.01.001>
- Regeer, B. J., & Bunders, J. G. F. F. G. (2009). Knowledge co-creation: Interaction between science and society. In *A Transdisciplinary Approach to Complex Societal Issues*. Den Haag: Advisory Council for Research on Spatial Planning, Nature and the Environment/Consultative Committee of Sector Councils in the Netherlands [RMNO/COS]. Advisory Council for Spatial Planning, Nature and the Environment (RMNO).
- Regioplan. (2022). *Leren en/of verantwoord? Ex durante evaluatie van de pilot Lerend Evalueren*. Amsterdam
- Rittel, H. W. J., & Webber, M. M. (1973). Dilemmas in general theory of planning. *Policy Sciences*, 4, 155–169.
- Roux, D. J., Stirzaker, R. J., Breen, C. M., Lefroy, E. C., & Cresswell, H. P. (2010). Framework for participative reflection on the accomplishment of transdisciplinary research programs. *Environmental Science & Policy*, 13(8), 733–741. <https://doi.org/10.1016/j.envsci.2010.08.002>
- Sarewitz, D. (2004). How science makes environmental controversies worse. *Environmental Science & Policy*, 7(5), 385–403. <https://doi.org/10.1016/j.envsci.2004.06.001>
- Schmidt, R. (2016). The Methodological Challenges of Practising Praxeology. In M. Lamers, G. Spaargaren, & D. Weenink (Eds.), *Practice Theory and Research: Exploring the Dynamics of Social Life* (pp. 59–75). Routledge.

<https://doi.org/10.4324/978131565690-13>

- Scholz, R. W., & Steiner, G. (2015). The real type and ideal type of transdisciplinary processes: part I—theoretical foundations. *Sustainability Science*, 10(4). <https://doi.org/10.1007/s11625-015-0326-4>
- Schulz, M., Kunseler, E.-M., Ophoff, P., Van der Steen, M. (2020). *Leren institutionaliseren - Reflecties bij het leren door de Rijksoverheid in de deal-aanpak*. NSOB en PBL, Den Haag
- Stake, R. E., & Abma, T. A. (2005). Responsive Evaluation. In S. Mathison (Ed.), *Encyclopedia of Evaluation*.
- Van der Steen, M., Faber, A., Frankowski, A., & Norbruis, F. (2018). *Opgavegericht evalueren: Beleidsevaluatie voor systeemverandering*. [https://www.nsob.nl/wp-content/uploads/2018/12/NSOB-18-17-Essay-OpgavegerichtEvalueren\\_web\\_20181214-2.pdf](https://www.nsob.nl/wp-content/uploads/2018/12/NSOB-18-17-Essay-OpgavegerichtEvalueren_web_20181214-2.pdf)
- Van Der Steen, M., Scherpenisse, J., Van Gerwen, O. J., & Kruitwagen, S. (2014). *Leren door doen - Overheidsparticipatie in een energieke samenleving*.
- Van Mierlo, B., Regeer, B. J., van Amstel, M., Arkesteijn, M., Beekman, V., Bunders, J., de Cock Buning, T., Elzen, B., Hoes, A.-C., & Leeuwis, C. (2010). *Reflexive monitoring in action: A guide for monitoring system innovation projects*.
- Van Veen, S. C., Verwoerd, L., & Regeer, B. J. (2016). *Characteristics of reflexive evaluation - a literature review conducted in the context of the Natuurpact (2014-2027) evaluation*. Athena Instituut, VU University Amsterdam, Amsterdam.
- Verwoerd, L., Klaassen, P., & Regeer, B. J. (2019). *The value of a reflexive evaluation approach in the eyes of researchers*. Athena Instituut, VU University Amsterdam, Amsterdam.
- Verwoerd, L., Klaassen, P., & Regeer, B. J. (2021). How to normalize reflexive evaluation? Navigating between legitimacy and integrity. *Evaluation*, 27(2), 229–250. <https://doi.org/10.1177/1356389020969721>
- 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 and Policy*, 103, 32–40. <https://doi.org/10.1016/j.envsci.2019.10.011>
- Verwoerd, . (2022). *Towards a practice of reflexivity: on the normalisation of knowledge co-production in policy evaluation at the PBL Netherlands Environmental Policy Assessment Agency*. Athena Instituut, VU University Amsterdam, Amsterdam.
- Weterings, A., & Verwoerd, L. (2020). *Verkennen waar wat werkt. Vormgeven aan lerend en adaptief beleid*. PBL, Den Haag