

The value of reflexive evaluation – a review of the Natuurpact evaluation (2014-2017)



Athena Institute *for Research on
Innovation and Communication in Health
and Life Sciences*

The value of reflexive evaluation – a review of the Natuurpact evaluation (2014-2017)

L. Verwoerd

R. De Wildt-Liesveld

B.J. Regeer



Athena Institute *for Research on Innovation and
Communication in Health and Life Sciences*

The value of reflexive evaluation – a review of the Natuurpact evaluation (2014-2017)

Athena Institute, VU University Amsterdam, commissioned by the PBL Environmental Assessment Agency

External reviewers

Prof. Dr. Jurien Edelenbos (EUR)

Prof. Dr. Hens Runhaar (WUR)

Prof. Dr. Bas Arts (WUR)

Internal reviewers

Rob Folkert (projectleader Natuurpact reflexive evaluation, PBL)

Froukje Boonstra (acting projectleader Natuurpact reflexive evaluation, WUR)

Femke Verwest (supervisor Natuurpact reflexive evaluation, PBL)

Ton Manders (chief scientist, PBL)

Eva Kunseler (project advisor, PBL)

Contact

Lisa Verwoerd (l.verwoerd@vu.nl), Barbara Regeer (b.j.regeer@vu.nl)

Authors

Lisa Verwoerd, Renée de Wildt-Liesveld en Barbara Regeer (all Athena Institute, VU University Amsterdam)

With special regards to:

All respondents (national and provincial policymakers, representatives from societal organisations, evaluation researchers) for taking part in our interviews.

All nature pictures in this report are courtesy of *Unsplash*: <https://unsplash.com/>

The Athena Institute

The Athena Institute is a research and education department at the Faculty of Beta Sciences, VU Amsterdam. The Athena Institute develops and applies transdisciplinary approaches for joint reflection and action on innovation processes among a wide variety of stakeholders who have different interests, values, visions in order to address complex problems in stagnating systems (society or organizations) in a sustainable and equitable way.

July 2017

The authors bear full responsibility for the content of the review but accept no liability whatsoever for any damages that may occur as a consequence of using the information in this publication. The Creative Commons license 4.0 for the Netherlands (attribution, non-commercial, no derivative works) is applicable to this work. *The value of reflexive evaluation – review of the Natuurpact evaluation (2014-2017)* by Lisa Verwoerd, Renée de Wildt-Liesveld & Barbara Regeer, is licensed under a creative commons attribution-noncommercial-noderivatives 4.0 INTERNATIONAL LICENSE

<https://creativecommons.org/licenses/by-nc-nd/4.0>



Table of Contents

1. Introduction.....	ii
1.1 Origin of the Natuurpact agreement and its evaluation.....	1
1.2 Call for reflexive evaluation.....	2
1.3 Review of the Natuurpact reflexive evaluation.....	3
1.4 Reader's guide.....	5
2. The Natuurpact evaluation.....	7
2.1 Setting the scene: the context in which the evaluation took place.....	7
2.2 Actors involved with the evaluation.....	8
2.3 The Natuurpact evaluation scope.....	11
2.4 The evaluation design.....	12
3. Theoretical background.....	19
3.1 The expected outcomes of reflexive evaluation.....	19
3.2 Conditional factors to realise the expected outcomes.....	22
4. Methodology.....	25
4.1 Data collection and analysis.....	25
4.2 Learning questions analysis.....	28
4.3 Research reliability and validity.....	29
5. The value of the Natuurpact evaluation.....	30
5.1 Five different value-types.....	30
5.2 Remaining informational needs.....	39
5.3 Reflection on the manifested values and remaining informational needs: knowledge-enriched policy practice.....	43
5.4 Reflection on theoretical framework in relation to findings.....	44
6. Observed factors that contributed to value.....	45
6.1 Stakeholder selection.....	46
6.2 Stakeholder engagement.....	48
6.3 Alignment between policy practice and evaluation research.....	55
6.4 Factors contributing to strategic value.....	61
6.5 Reflection on theoretical framework in relation to our findings.....	62
7. Final conclusions and recommendations.....	64
7.1 Conclusions.....	64
7.2 Discussion.....	66
7.3 Recommendations for the continuation of the evaluation.....	Error! Bookmark not defined.
References.....	72

Nederlandse samenvatting

1. Aanleiding review lerende evaluatie Natuurpact

In de afgelopen jaren is het natuurbeleid in Nederland gedecentraliseerd. Het Natuurpact (2013) vormt het bestuurlijk sluitstuk op dit decentralisatieproces. De twaalf Nederlandse provincies maken én implementeren beleid om samen met het Rijk de gestelde doelen te realiseren in 2027. Volgens het Natuurpact richten de provincies en Rijk zich op de realisatie van het Natuurnetwerk Nederland, het halen van de internationale doelen (de Vogel en Habitat richtlijn, VHR, en de Kaderrichtlijn Water, KRW), het versterken van de betrokkenheid van de samenleving bij de natuur en de verbinding van economie en natuur. Het Ministerie van Economische Zaken (EZ) en het Interprovinciaal Overleg (IPO) – als vertegenwoordiger van de provincies – hebben het Planbureau voor de Leefomgeving (PBL) gevraagd om eens in de drie jaar te evalueren hoe de afspraken uit het Natuurpact vorderen. PBL voert deze evaluatie uit in samenwerking met partner WUR. De eerste evaluatie periode heeft plaatsgevonden in de periode 2015 tot 2017 en heeft geresulteerd in het hoofdrapport *Lerende evaluatie van het Natuurpact*, drie achtergrondrapporten (*De praktijk van vernieuwingen in het provinciaal natuurbeleid; Het provinciaal natuurbeleid ingekaderd en Potentiele bijdrage provinciaal natuurbeleid aan Europese biodiversiteitsdoelen*) en een themasite waar deze rapporten en meer achtergrond informatie over de evaluatie te vinden is¹.

De bestuurlijke context waarin natuurbeleid ontwikkeld en uitgevoerd wordt is de afgelopen jaren veranderd. Met de decentralisatie zijn bevoegdheden en verantwoordelijkheden verschoven van het Rijk naar de provincies (multi-level). Daarnaast wordt beleid door verschillende overheden steeds nadrukkelijker samen met andere partijen in het veld vormgegeven en uitgevoerd (multi-actor). Bovendien is de liggende opgave voor de VHR en KRW groot én is de ambitie verbreed: niet alleen biodiversiteit, maar het vergroten van maatschappelijke betrokkenheid en het versterken van de verbinding tussen natuur en economie zijn doelen waar natuurbeleid naar streeft. Deze ontwikkelingen – de veranderde bestuurlijke context, de grote opgave voor de VHR en de KRW én de verbrede ambitie – zijn grote veranderingen waarbij nog veel te leren valt hoe hier goed invulling aan te geven. Vanwege deze transities hebben de opdrachtgevers samen met PBL geen reguliere impact evaluatie uit te voeren, maar een *lerende evaluatie*.

Lerend evalueren is een relatief nieuwe methode – ook voor het PBL. De onderzoekers stappen daarmee van hun rol als afstandelijke onderzoekers en zijn tijdens het evaluatie proces meer interactief betrokken geweest bij de provincies wiens beleid geëvalueerd is en andere maatschappelijke partijen die daarmee gemoeid zijn. Ook voor opdrachtgevers en partijen die actief bij de evaluatie zijn betrokken (met name de provincies, maar ook het Rijk, maatschappelijke organisaties, bedrijven) betekende deze evaluatie een andere rol; hen werd gevraagd actiever betrokken te zijn bij alle fasen van het evaluatieonderzoek.

¹ Alle rapporten zijn te downloaden via: <http://themasites.pbl.nl/evaluatie-natuurpact/>

Vanwege de nieuwe aanpak heeft PBL het Athena Instituut van de Vrije Universiteit (vanaf hier: de auteurs) gevraagd om het evaluatieproces mede te ontwerpen en faciliteren en de opbrengsten van de lerende evaluatie te evalueren. Onderdeel hiervan was een wetenschappelijke onderbouwing van het procesontwerp bestaande uit een theoretisch kader met daarin de belangrijkste karakteristieken van een lerende evaluatie (*Characteristics of reflexive evaluation – a literature review conducted in the context of the Natuurpact evaluation (2014-2017)*)². In dit huidige rapport gebruiken we dit theoretisch kader om te reflecteren op de toepassing van lerend evalueren zoals uitgevoerd door PBL en de WUR. PBL en de WUR veronderstelden bij aanvang van de evaluatie dat lerend evalueren de kwaliteit van kennis, de bruikbaarheid en daardoor ook de impact van het onderzoek zou vergroten. In deze review geven we ten eerste inzicht in wat de aanpak concreet heeft opgeleverd om te zien of de veronderstelde waarden van lerend evalueren daadwerkelijk zijn geogst en ten tweede welke aanpassingen het proces verder kunnen verbeteren. We beantwoorden daarmee de volgende twee hoofdvragen:

- I. Op welke wijze is de lerende evaluatie van het Natuurpact van waarde geweest voor het Rijk en de provincies en hoe verhouden deze waarden zich tot de beoogde waarden? (zie sectie 3)
- II. Hoe heeft sturing op basis van het theoretisch kader (Van Veen et al., 2016) bijgedragen aan het realiseren van waarde en welke andere factoren kunnen we onderscheiden die hebben bijgedragen aan de waarde van de evaluatie? (zie sectie 4)

Om deze vragen te beantwoorden hebben we de perceptie van waarden door beleidsbetrokkenen die deel hebben genomen aan de evaluatie in kaart gebracht. Onze data is gebaseerd op participatieve observatie gedurende de evaluatieperiode, de analyse van audio-opnamen, transcripten en samenvattingen van workshops en interviews met beleidsbetrokkenen waarbij wij niet aanwezig waren, 11 semi-gestructureerde interviews met deelnemers en onderzoekers en een focusgroep discussie met leden van de IPO werkgroep Natuurbeleid. Deze werkgroep bestaat uit één vertegenwoordiger uit elke provincie, een voorzitter (uit provincie Gelderland) en een secretaris vanuit het IPO. Data is geanalyseerd aan de hand van concepten uit het theoretisch kader. Daarnaast is er een analyse gedaan van leervragen – gearticuleerd tijdens de evaluatie periode – op basis van alle beschikbare transcripten. We hebben ons tijdens dit onderzoek beperkt tot de ervaringen en percepties van de beleidsbetrokkenen van de evaluatie. In een vervolgstudie brengen we de verschillende wijzen waarop de lerende evaluatie van waarde is geweest voor de evaluatieonderzoekers in kaart.

2. De Natuurpact lerende evaluatie

Het theoretisch kader

Voor we onze resultaten bespreken geven we eerst een korte beschrijving van a) het theoretisch kader naar Van Veen et al. (2016), op basis waarvan de Natuurpact lerende evaluatie is vormgegeven, en b) de daadwerkelijke uitvoering van de evaluatie. Figuur 1 geeft het theoretisch kader weer, met

² Deze literatuur review is tevens te downloaden via: <http://themasites.pbl.nl/evaluatie-natuurpact/wp-content/uploads/characteristics-of-reflexive-evaluation.pdf>

links de factoren die van belang zijn bij het vormgeven van een lerende evaluatie, en rechts de uitkomsten die worden verwacht van een dergelijk proces.

Om te beginnen met de uitkomsten (zie rechterhelft van figuur 1); de evaluatieonderzoekers verwachtten dat de evaluatie kennis zou produceren die sociaal robuust is (betrouwbaar, relevant en toepasbaar). Omdat deze kennis gezamenlijk met de betrokkenen wordt ontwikkeld, kunnen inzichten direct gebruikt worden om het beleid en/of de uitvoeringspraktijk tussentijds aan te passen. De verwachting is daarom dat een lerende evaluatie leidt tot bruikbaarere kennis, en daardoor *kennis-verrijkte beleidspraktijken* en uiteindelijk tot meer *beleidsimpact*. Daarnaast beoogden de onderzoekers middels een lerende evaluatie een brug te slaan tussen twee belangrijke functies van evalueren: *verantwoorden* van beleid (niet alleen 'omhoog', richting opdrachtgevers, maar juist horizontaal: richting alle partijen die betrokken zijn bij het beleid), en *leren* om beleid (de vormgeving en de uitvoering) te verbeteren.

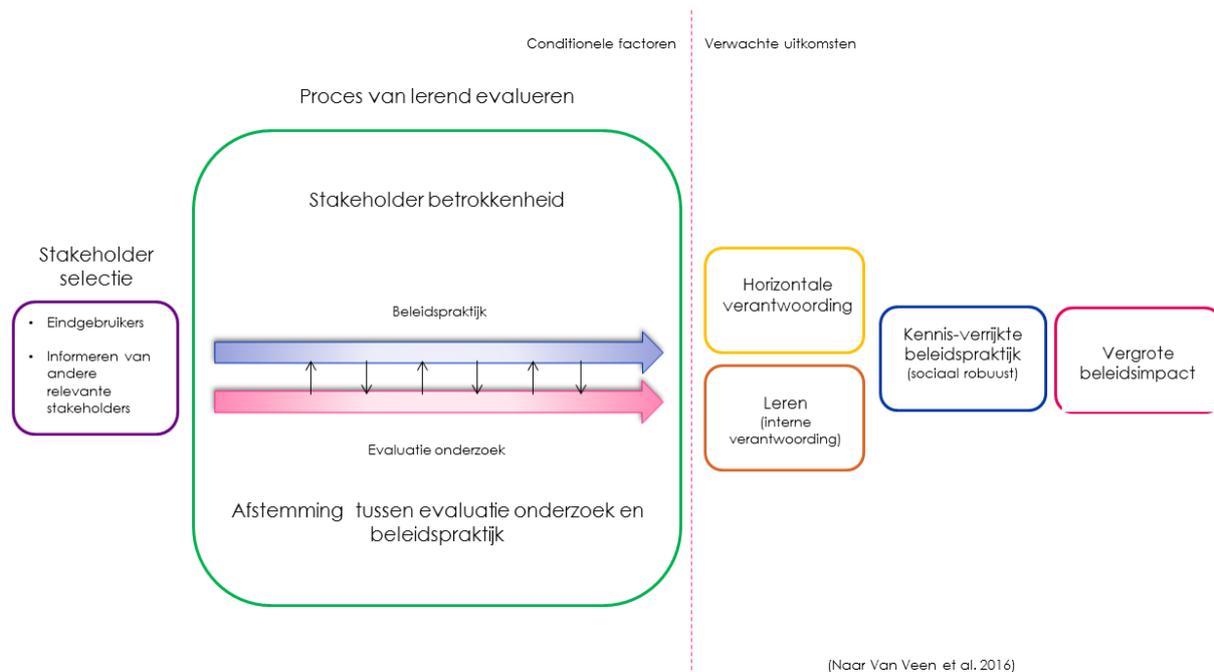
Wat is er voor nodig om deze uitkomsten te realiseren? Onder conditionele factoren vallen, naast het selecteren van de relevante stakeholders, het realiseren van stakeholder betrokkenheid en het

organiseren van continue afstemming tussen het evaluatieonderzoek en de geëvalueerde beleidspraktijk (zie linkerhelft Figuur 1). Met betrekking tot het selecteren van relevante stakeholders zegt het theoretisch kader dat de primair te betrekken groep de eindgebruikers van de evaluatie zouden moeten zijn. Daarnaast is het van belang ook andere actoren geïnformeerd te houden en een goede relatie te onderhouden met de



Schotse Hooglander op de Oostvaardersplassen in Flevoland

opdrachtgevers om politiek-bestuurlijke steun voor de lerende evaluatie te garanderen. Met betrekking tot het realiseren van stakeholder betrokkenheid dient er geïnvesteerd te worden in het creëren van betrokkenheid en eigenaarschap bij de eindgebruikers, door het creëren van een transparant onderzoeksproces en regelmatige communicatie hierover, waardoor wederzijds vertrouwen kan ontstaan. Continue afstemming tussen het evaluatieonderzoek en de beleidspraktijk wordt bereikt door het betrekken van de stakeholders in iedere fase van het onderzoek en het aanpassen van het onderzoeksontwerp en -proces naar aanleiding van stakeholder-input. Hiervoor is regelmatige interactie van belang, waarin wordt gereflecteerd op verschillende aspecten van de evaluatie en het leerproces wordt gestimuleerd.



Figuur 1. Theoretisch kader van lerend evalueren naar Van Veen et al. 2016.

Het verloop van de Natuurpact evaluatie

De lerende evaluatie van het Natuurpact kan grofweg onderscheiden worden in vier opeenvolgende fasen: het bepalen van de evaluatie-doelen en aanpak, de data verzameling, de data analyse en interpretatie en het formuleren van handelingsperspectieven en de verspreiding van het eindrapport. Tijdens elke fase zijn diverse workshops en andere interactie momenten geweest tussen de deelnemers van de evaluatie (met name de provincies, vertegenwoordigers van het Rijk en maatschappelijke partijen). Figuur 2 is een schematische weergave van het verloop van de evaluatie.

Tabel 1. Een overzicht van de evaluatieactiviteiten en betrokken actoren per fase.

Evaluatie fasen	0. Voor onderzoek			1. Bepalen evaluatie doelen en aanpak			2. Data verzameling			3. Data analyse en interpretatie			4. Formuleren handelingsperspectief en en verspreiding eindrapport	
	Vooronderzoek: interviews	Leersessie I	Leersessie II	Leersessie III	Interviews	Bilaterale consultatie	Case studies	Werkessie Kosteneffectiviteit	Werkessie Provinciale beleidsvernieuwingen	Werkessie Handelingsperspectieven	Groeps-review sessies	Expert sessie Handelingsperspectieven	Gevraagd voor feedback op rapport	
Provinciale beleidsmedewerkers	x	x	x	x	x	x	x	x	x	x	x	x	x	
Opdrachtgevers (EZ en IPO)	x	x	x	x					x	x	x	x	x	
Maatschappelijk partners	x		x	x	x		x		x	x				

Hoewel we ons realiseren dat we niet volledig zijn, geeft Tabel 1 de voornaamste evaluatie activiteiten weer waar onderzoekers en deelnemers aan de evaluatie bijeenkwamen om de evaluatie vorm te geven en uit te voeren. Een belangrijke toevoeging op de tabel zijn de overleggen met het BACVP (Bestuurlijke Advies Commissie Vitaal Platteland, bestaande uit 12 Gedeputeerden natuur) en het Informeel Bestuurlijk Overleg (waarbij naast de 12 Gedeputeerden ook de staatssecretaris van EZ aanwezig was) waarbij afstemming over de evaluatie plaatsvond, evenals het delen en interpreteren van de eerste evaluatie bevindingen. De eerste evaluatierapportage is afgerond in januari 2017.

3. De waarde van lerend evalueren

Welke waarde heeft de evaluatie volgens betrokkenen gehad?

De evaluatie is op verschillende manieren van waarde geweest voor de betrokkenen. We onderscheiden inhoudelijke waarde, affectieve waarde, strategische waarde, netwerk waarde en instrumentele waarde.

De betrokkenen geven ten eerste aan dat de evaluatie van *inhoudelijke waarde* is geweest voor hen; de evaluatie heeft bijgedragen aan kennis-verrijkte beleidspraktijken. In de eerste plaats heeft de evaluatie kennis opgeleverd over de stand van zaken met betrekking tot de potenties van het provinciale beleid voor het bereiken van de doelen, met name als het gaat om internationale biodiversiteitsdoelen. Met betrekking tot de andere twee ambities, maatschappelijke betrokkenheid en natuur en economie, is vooral helder geworden dat men belang hecht aan verdere uitwerking, maar dit is in deze evaluatieperiode nog nauwelijks gebeurd. In de tweede plaats heeft de evaluatie kennis opgeleverd over de verschillende mogelijke strategieën om beleidsdoelen te realiseren, en is hiermee transformationele kennis ontwikkeld. In dit kader heeft men ook geleerd over de eigen en elkaars denk- en handelingskaders en hoe die de beleidspraktijk beïnvloeden. Ten slotte zien we dat, als gevolg van het gezamenlijke leerproces, er zich een gezamenlijke taal en gedeelde ambitie heeft ontwikkeld.

Ten tweede hechten de betrokkenen veel belang aan de *affectieve waarde* die de evaluatie heeft opgeleverd. Het samen leren van elkaars beleidspraktijk heeft een gevoel van saamhorigheid en onderling vertrouwen opgeleverd. Het bespreken van kansen maar ook van uitdagingen heeft ook gezorgd voor een gevoel van geruststelling ('wij zijn niet de enige die hier mee worstelen').

Naast inhoudelijke en affectieve waarde heeft de evaluatie, ten derde, *strategische waarde* gehad. Het gaat hierbij om het legitimeren van bepaalde keuzes (aan collega's en maatschappelijke organisaties) die provinciale beleidsmakers maken, met name met betrekking tot de verbrede ambities ('het staat in het PBL rapport, dus is het een juiste beslissing'). We zien inderdaad dat de beleidsdiscourse is opgeschoven naar een acceptatie van de verbrede ambities (naast biodiversiteitsdoelen, ook doelen op het gebied van natuur en samenleving en natuur en economie). Bovendien heeft de evaluatie een gevoel van urgentie gegeven aan de verbreding, en deze op de agenda gezet in verschillende provincies. Ten slotte zien we dat (hoewel het nadrukkelijk niet de intentie was van de evaluatie) de evaluatie op hogere overheidsniveaus bijdraagt aan de legitimering van de decentralisatie van natuurbeleid, middels de verwachte successen die kunnen worden geboekt met betrekking tot de Nederlandse biodiversiteit. Mogelijk draagt dit zelfs bij aan de legitimering van de verbreding van natuurbeleid op deze zelfde overheidsniveaus.

Ten vierde heeft de evaluatie *netwerk waarde* gehad voor de betrokkenen. Beleidsbetrokkenen hebben andere relevante actoren leren kennen en daarmee hun netwerk uitgebreid. Verder geven provinciale beleidsmedewerkers aan dat de onderlinge relaties zijn versterkt en dat de relatie met het ministerie van EZ is verbeterd als gevolg van het evaluatieproces.

Ten slotte heeft de evaluatie, in mindere mate, *instrumentele waarde* gehad. Instrumentele waarde is de vertaling van inhoudelijke waarde naar concrete acties en beslissingen. Het is op dit moment nog te vroeg om te kunnen zien hoe de eerste evaluatie van het Natuurpact doorwerkt in de beleidspraktijk. We hebben wel gezien dat provincies zich hebben laten inspireren door de evaluatie bij het formuleren van hun Natuurvisies en de staatssecretaris van EZ heeft in een formele brief, als reactie op vragen van de Eerste en Tweede Kamer, aangegeven dat Rijk en provincies met een plan van aanpak zullen komen waarin zij aangeven hoe ze om zullen gaan met de aanbevelingen uit het evaluatierapport.

De netwerk waarde en affectieve waarde zijn specifiek voor de lerende evaluatie. De lerende evaluatie heeft specifieke meerwaarde gehad door de combinatie van een doelevaluatie (doelbereik en efficiëntie) met een procesevaluatie (hoe worden deze doelen bereikt). Bij de procesevaluatie is deze evaluatie nog een stap verder gegaan door veel aandacht te besteden aan de vraag hoe deze doelen bereikt *kunnen* worden en door aandacht voor reflectie op eigen en andermans denk- en handelingskaders die het bereiken van deze doelen al dan niet dichterbij brengen. Dit maakt dat inhoudelijke en instrumentele waarde ook eerder worden herkend en geaccepteerd, wat kan leiden tot toepassing van de opgedane kennis in de beleidspraktijk. Daarnaast levert een lerende evaluatie, vanwege de interactie met de praktijk, nieuwe en andere kennis op, en daarmee een andere invulling van de inhoudelijke waarde dan bij een reguliere evaluatie. Vooral vanwege de toegenomen complexiteit van het natuurbeleid is interactie met andere partijen nodig om de kennis te verkrijgen die nodig is om het systeem te kunnen doorgronden.

Een andere opvallende waarneming is dat elk van de waarden is gemanifesteerd zowel op het niveau van de primaire eindgebruiker van de evaluatie (de provincies) en het collectief niveau van het gezamenlijk netwerk dat is ontstaan gedurende het evaluatieproces (zie Tabel 2). Deze onverwachte collectieve waarden zijn mogelijk karakteristiek voor de context zoals die van de Natuurpact evaluatie, namelijk die van een grootschalig, complex beleidsprogramma waarbij beleidsontwikkeling en uitvoering plaatsvindt op verschillende overheidsniveaus in verschillende regio's.

Tabel 2. Overzicht van de waarde die de evaluatie heeft gehad volgens betrokkenen, op twee niveaus (gearceerde roze vlakken geven aan waar de lerende evaluatie specifieke meerwaarde had ten opzichte van reguliere evaluatie)

	Netwerk waarde	Affectieve waarde	Inhoudelijke waarde	Instrumentele waarde	Strategische waarde
Provinciaal niveau	Netwerk uitbreiding	Opluchting en geruststelling Gevoel van saamhorigheid	Kennis over: <ul style="list-style-type: none"> Huidige stand van zaken (m.b.t. plan- en uitvoeringspotentieel) Variëteit aan strategieën Verheldering van: <ul style="list-style-type: none"> Onderliggende aannames Frame reflectie 	Aanvulling op en toevoeging aan Natuurvisies	Legitimatie verbrede ambitie Gevoel van urgentie Agendasetting
Collectief niveau	Versterkt interprofessionele relaties tussen verschillende overheden	Vertrouwen tussen verschillende overheden	Meer gedeelde conceptuele samenhang Gedeelde ambitie	Toegenomen samenhang van beleidsprocessen tussen verschillende overheden	Legitimatie van decentralisatie Mogelijk ook legitimatie van verbrede ambities op hogere overheidsniveau
Aan welke verwachte uitkomsten dragen waarden vooral bij?	Leren Horizontale verantwoording	Leren	Leren Kennis-verrijkte beleidspraktijken	Kennis-verrijkte beleidspraktijken Toegenomen beleidsimpact	Toegenomen beleidsimpact Horizontale verantwoording

Hoe verhouden deze waarden zich tot de verwachte uitkomsten?

De laatste rij in Tabel 2 geeft weer hoe de gevonden waarden zich verhouden tot de verwachte uitkomsten, zoals weergegeven in het theoretisch kader in Figuur 1. De verschillende waargenomen waarden, met name inhoudelijke en instrumentele, laten zien dat we inderdaad kunnen spreken van *kennis-verrijkte beleidspraktijken*. Wat betreft toegenomen *beleidsimpact* valt op dit moment nog weinig hard te maken; het is op dit moment nog te vroeg na de eerste evaluatie periode om in kaart te brengen hoe de evaluatie beleidsimpact heeft beïnvloed. Desalniettemin is al wel zichtbaar dat de aanbevelingen uit het eindrapport tot diverse acties aanzetten. Daarnaast zien we dat de evaluatie waarden heeft opgeleverd die niet geanticipeerd zijn, maar wel zeer gewaardeerd: de netwerk en affectieve waarde. Deze waarden hebben zich vertaald tot een *lerend beleidsnetwerk*. Naast het bevorderen van *leren* heeft de evaluatie ook de functie van *verantwoorden*. We hebben gezien dat verantwoord en snel wordt vertaald naar de angst te worden afgerekend op (tegenvallende) resultaten. Dit heeft het leren tot op zekere hoogte beperkt. Zo heeft men niet altijd het achterste van de tong laten zien in gezamenlijke leersessies. En, om een veilige situatie te creëren voor kennisuitwisseling tussen beleidsmakers van provincies is op sommige momenten de actorenselectie beperkt gebleven, wat kennisverrijking door betrokkenheid van een diverse groep stakeholders heeft

beperkt. Tegelijkertijd heeft deze angst voor afrekenen er ook voor gezorgd dat er een hoge opkomst en actieve betrokkenheid was van provinciale beleidsbetrokkenen bij gezamenlijke leersessies, wat het onderlinge leren heeft bevorderd. Een tweede voorbeeld waarbij verantwoord en leren elkaar in eerste instantie tegenwerkten, maar elkaar ook hebben versterkt heeft te maken met dataverstrekking door provincies. Het leveren van de benodigde gegevens door de provincies is, zeker in het begin, vrij moeizaam gegaan. De extra aandacht die er als gevolg hiervan is gegaan naar de provincies, middels bilaterale consultaties, heeft juist volgens de betrokkenen een groot aandeel gehad in het leren van de evaluatie. Wie zien dus dat de twee functies van de lerende evaluatie (verantwoorden en leren) elkaar zowel versterken als beperken.

Er zijn nog openstaande kennisbehoeften

Naast de verschillende wijzen waarop de evaluatie van waarde is geweest, zien we ook dat de evaluatie – in deze eerste periode – niet alle kennisbehoeften van de beleidsbetrokkenen voldoende heeft vervuld. Naast interviews met beleidsbetrokkenen hebben we op basis van participatieve observatie en documentanalyse een aantal terugkerende *leervragen* kunnen identificeren waar nog geen antwoord of handelingsperspectief voor is en waar een volgende evaluatie meer aandacht aan zou kunnen besteden.

Ten eerste zien we dat de bredere ambities van het Natuurpact (maatschappelijke betrokkenheid en verbinding natuur en economie) nog niet zijn vertaald naar concrete doelen en bijbehorende indicatoren. Hierbij is men bezorgd dat doelen te rigide geformuleerd zullen worden, waardoor ze niet effectief zijn voor het bereiken van meer maatschappelijke betrokkenheid en een verbinding tussen natuur en economie. Hoe behulpzame, niet-rigide doelen te ontwikkelen (die ruimte overlaten voor afweging en experimenteren) is niet helder; hier ligt een kennisbehoefte. Verder zien we dat direct betrokkenen de verbreding van de ambities van het Natuurpact hebben omarmd, maar hun collega's, maatschappelijke partners en relevante actoren in het bestuurlijke en politieke netwerk nog niet. Pioniers hebben handelingsperspectieven nodig om hier mee om te gaan.

Ten tweede liggen er vragen rondom de biodiversiteitsdoelen die momenteel voornamelijk worden ingevuld aan de hand van VHR-doelstellingen. Men ervaart dit als beperkend en niet altijd bijdragend aan het realiseren van biodiversiteit. Met name de focus op het realiseren van VHR-doelen in het Natuurnetwerk (en Natura 2000) roept de vraag op wat de rol van de gebieden eromheen is. Sommigen geven aan dat de volgende evaluatie zich niet zou moeten beperken tot het beleid dat in het Natuurpact omschreven staat, maar ook bijvoorbeeld het Rijksbeleid voor natuur en het waterbeleid moet omvatten. Deze meer integrale benadering, waarbij wordt gestart met biodiversiteit in de Nederlandse natuur, en niet met bestaande beleidskaders, roept nieuwe kennisvragen op en heeft ook consequenties voor de te betrekken actoren bij de volgende evaluatie.

Ten derde hebben we veel vragen gezien rondom de nieuwe rol van provincies bij het realiseren van meer maatschappelijke betrokkenheid bij natuurbeleid. Dit levert dilemma's op zoals, hoe kunnen we anderen meer eigenaarschap geven terwijl wij als provincies wel onze eigen doelstellingen hebben, en hoe gaan we om met verschillende, soms tegengestelde, belangen in een gebied? Het versterken van het leervermogen van provincies op dit gebied is gewenst, en wetenschappelijke kennis over het faciliteren van multi-stakeholderprocessen en de rolverschuiving van een presterende, rechtmatige overheid naar een participerende, faciliterende overheid kunnen hierbij een rol spelen.

4. Factoren die hebben bijgedragen aan waarde

Op basis van ons theoretisch kader onderscheidde we drie categorieën factoren die hebben bijgedragen aan de waarde van de evaluatie: stakeholder selectie, stakeholder betrokkenheid en afstemming tussen beleidspraktijk en evaluatieonderzoek. Deze laatste categorie bevat de factoren die bepalend zijn geweest voor het realiseren van de geobserveerde waarden, en bespreken we daarom eerst.

Afstemming tussen beleidspraktijk en evaluatieonderzoek

Ten eerste, met betrekking tot de inhoudelijke waarde van de evaluatie, hebben de onderzoekers actief gestuurd op afstemming op de behoeften uit de beleidspraktijk. Provincies zijn betrokken geweest tijdens de meeste evaluatiefasen en hebben bijvoorbeeld meegedacht en hun akkoord gegeven over de inhoud van de evaluatie en de inhoud van de werksessies. Echter, de rol van provincies is nog redelijk beperkt gebleven en veel van de inhoudelijke controle lag bij de onderzoekers. Zo hebben provincies bijvoorbeeld geen input gegeven op de keuzes voor evaluatiemethoden. Dit had enerzijds pragmatische overwegingen, anderzijds was het niet altijd makkelijk provincies te bewegen een actievere rol aan te nemen. Het is aannemelijk dat de inhoudelijke waarde van de evaluatie groter had kunnen zijn, hadden de provincies meer zeggenschap gehad. Om de afstemming op beleidspraktijk verder te vergroten, hadden onderzoekers veel aandacht voor de individualiteit van elke provincie. De onderzoekers' sensitiviteit voor de omstandigheden en behoeften van elke provincie was daarbij een belangrijke factor. Sensitiviteit wordt wel genoemd in literatuur, maar bleek in de praktijk van groter belang dan in eerste instantie verwacht en lijkt intuïtief aangevoeld te zijn door de onderzoekers. Dit kwam met name tot uiting tijdens de bilaterale consultaties (waarbij de onderzoekers elke provincie bezochten om data te verzamelen voor de evaluatie) – die werden zeer gewaardeerd door de gedetailleerde inhoudelijke afstemming waardoor de juiste data (en de juiste interpretatie en *framing* daarvan) overlegd kon worden en de uiteindelijke resultaten voor provincies herkenbaar waren. De consultaties waren van tevoren niet gepland, maar waren – hoewel tijdrovend – een belangrijke bepalende factor voor het inhoudelijke waarde én vertrouwen in de onderzoekers. Tevens kwam sensitiviteit tot uiting tijdens de groepsreview (waarin de voorlopige bevindingen van de biodiversiteitsanalyse en bijbehorende beleidsstrategieën werden gepresenteerd aan groepen van drie provincies per keer). Dat onderzoekers hier voorlopige resultaten deelden en zich kwetsbaar opstelden heeft in belangrijke mate bijgedragen aan het vertrouwen en de transparantie van de evaluatie.

Ten tweede bleek voor de affectieve waarde het ontmoeten van andere professionals die als het ware 'in hetzelfde schuitje verkeren' een bepalende factor. Kunnen spreken met gelijkgestemden en ervaringen en ideeën uit kunnen wisselen tijdens de diverse workshops droeg bij aan een gevoel van saamhorigheid en geruststelling. De groepsreview sessies speelden hier wederom een belangrijke rol, met name door de veilige context waarin oprechte ervaringen gedeeld konden worden. Hier is actief op gestuurd door in de vormgeving van interactiemomenten expliciet het delen van uitdagingen, worstelingen en onzekerheden op te nemen.

De strategische waarde – ten derde – heeft met name geprofiteerd van de onafhankelijke status van PBL en het eindrapport. De onderzoekers hebben hun onafhankelijkheid en daarmee geloofwaardigheid bewaakt door diverse strategieën (e.g. meer interactieve en meer afstandelijke rollen zijn verdeeld binnen het onderzoeksteam, interne review bij PBL, gewerkt met externe

wetenschappelijke reviewers). Dat de aanbevelingen onomstotelijk ‘van PBL’ zijn, maakte dat provincies deze konden inzetten om hun beleidskeuzes mee te legitimeren. Als er twijfel had bestaan over PBL’s onafhankelijkheid – bijvoorbeeld als de provincies coauteur waren geweest van het eindrapport, conform beschrijvingen van lerende evaluaties in de literatuur – had dit de strategische waarde van de evaluatie grotendeels teniet gedaan. Het bewaken van een zekere afstand, onder andere met betrekking tot het eindrapport, blijkt dus in de specifieke situatie van de Natuurpact evaluatie een belangrijke factor te zijn geweest, verbonden met de strategische waarde. Het hiermee gepaard gaande risico, van gebrek aan eigenaarschap over de bevindingen en aanbevelingen, is uitgebleven, dankzij de intensieve samenwerking in de fases van data verzamelen en analyse. Sommige provincies zijn kritisch op de *framing* van de bevindingen in het eindrapport: deze had scherper (‘minder lief’) mogen zijn, zodat deze meer urgentie had gegeven aan de verbreding. In hun perspectief had de strategische waarde van de evaluatie groter kunnen zijn. Vanwege de wens weg te blijven van het afrekenen van beleid, en juist het leren van de evaluatie centraal te zetten hebben onderzoekers hun bevindingen positief geformuleerd. Een ‘positieve evaluatie’ wordt in literatuur herkend als waardevol om het leren van evaluatie te vergroten – de conceptuele waarde van de evaluatie heeft in dit aspect voorrang gekregen op de strategische waarde.

Ten vierde profiteerde de netwerk waarde van de vele georganiseerde workshops waar verschillende partijen samenkwamen en waar inspirerende verhalen werden verteld over nieuwe samenwerkingen. Hier was in de plan-fase reeds op geanticipeerd en is expliciet op gestuurd.

Ten slotte onderscheiden we voor instrumentele waarde dezelfde factoren als voor de inhoudelijke waarde (waarvan instrumentele waarde in het verlengde ligt) en voegen we daar de factor timing aan toe. De opdrachtgevers hebben strak gestuurd op tijdige oplevering van resultaten. Interessant was dat timing niet alleen de verantwoordelijkheid was van de evaluatoren - sommige provincies gaven aan dat ze bij de planning van het ontwikkelen van de natuurvisie rekening hielden met de verschijningsdatum van het eindrapport.

Stakeholder selectie

Om afstemming tussen beleidspraktijk en evaluatieonderzoek te bereiken was dus regelmatige interactie tussen evaluatoren en stakeholders van groot belang. Hoe deze stakeholders zijn geselecteerd, en hun betrokkenheid gestimuleerd, zullen we hieronder toelichten.



Boerenwaluw

Aangezien de evaluatie plaatsvond in een context waarin natuurbeleid is gedecentraliseerd naar de provincies, zijn zij door de onderzoekers aangewezen als de voornaamste deelnemers – de primaire eindgebruikers – van de evaluatie (het proces en het eindrapport). De invulling van eindgebruikers is daarmee smaller dan in de literatuur bedoeld. Het Rijk is beperkter betrokken geweest – als opdrachtgever en

daarnaast met name als observant tijdens collectieve sessies – en daardoor zijn behoeften van het Rijk minder aan bod gekomen gedurende deze evaluatie periode, zoals leerbehoeften ten aanzien van haar nieuwe rol als systeemverantwoordelijke. Eveneens is de betrokkenheid van maatschappelijke actoren nog redelijk beperkt geweest; zij hebben kennis geleverd en hebben meegedacht over handelingsperspectieven naar aanleiding van de evaluatie resultaten, maar hun (leer)behoeften zijn niet meegenomen in de evaluatie. In wetenschappelijke literatuur wordt echter aanbevolen een brede selectie aan stakeholders te betrekken om zo gebruik te maken van zoveel mogelijk verschillende kennis. Bovendien wordt zo gezorgd dat alle relevante actoren deel uitmaakt van de ontwikkeling van nieuwe kennis én wordt voorkomen dat partijen zich buitengesloten en niet gehoord voelen, en daardoor gaandeweg dwars gaan liggen. In de context van de verbrede ambities van het Natuurpact, waarbij een toenemend aantal maatschappelijke actoren betrokken is, van waterschappen tot LTO en Natuurmonumenten, zou een bredere invulling van het begrip ‘eindgebruikers’ voor de hand liggen.

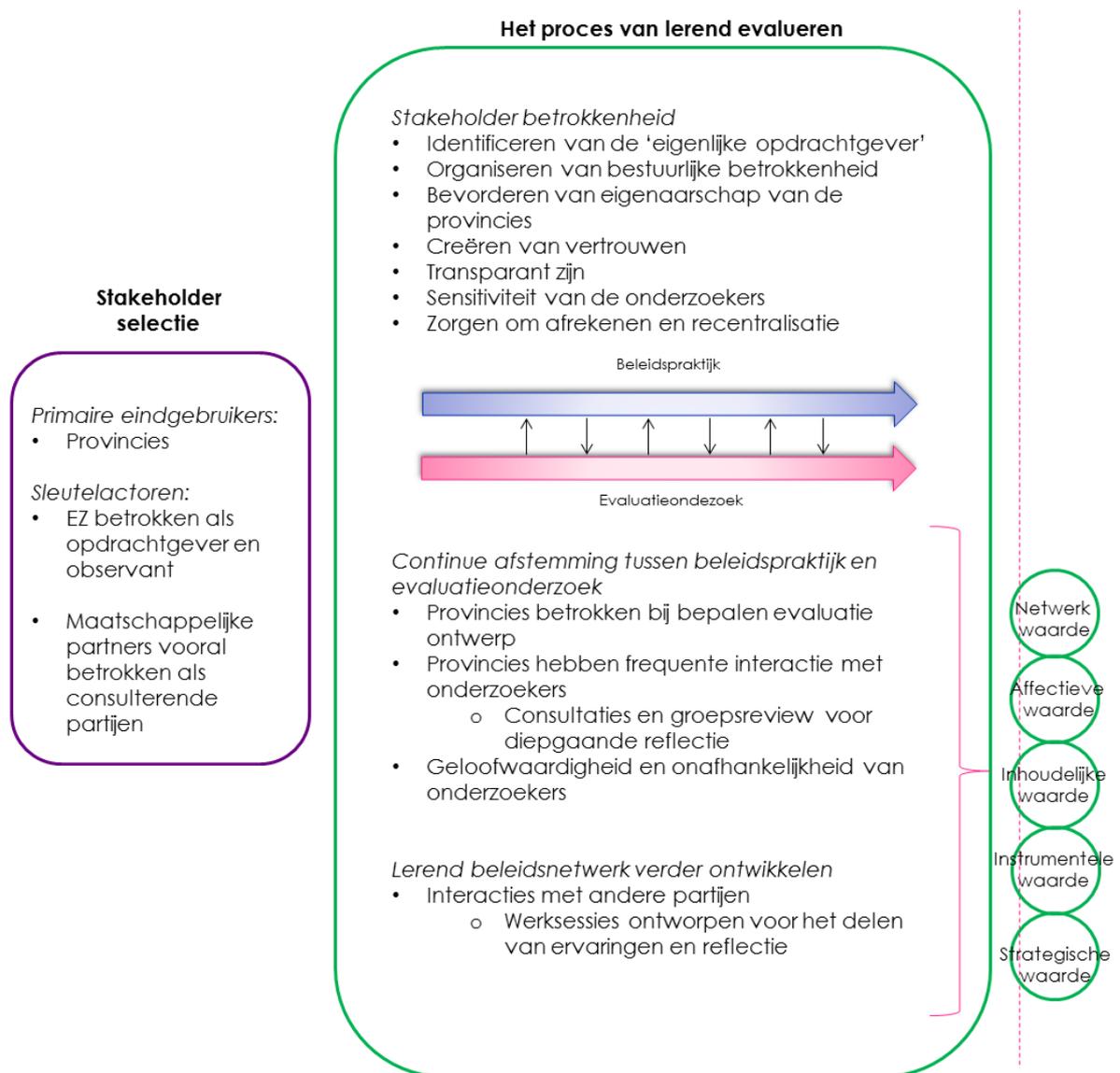
Belangrijk is om op te merken dat de gevoelige bestuurlijke context (de gespannen verhoudingen tussen Rijk en provincies) niet toestond een bredere stakeholderselectie aan te houden tijdens deze eerste evaluatieperiode. De provincies hun ervaringen laten delen met PBL – door sommigen van de provincies waargenomen als verlengstuk van het Rijk – en elkaar was al vooruitstrevend en vereiste een veilige omgeving. Het vergaand betrekken van Rijk en maatschappelijke partijen had deze veilige omgeving teniet gedaan, en had mogelijk deelname van de provincies ontmoedigd.

Stakeholder betrokkenheid

De gevoelige bestuurlijke context had ook invloed op de betrokkenheid van de provincies: deze was niet vanzelfsprekend. De provincies en het Rijk wantrouwden elkaar, wat maakte dat de provincies niet bepaald stonden te springen om deel te nemen aan de evaluatie, zeker daar een aantal van hen het PBL zien als een verlengstuk van het Rijk. De provincies waren bezorgd dat de evaluatie met name gebruikt zou worden door het Rijk om recentralisatie van het beleid te legitimeren.

De onderzoekers hebben verschillende strategieën toegepast om betrokkenheid van de provincies te stimuleren. Eén zeer belangrijke strategie begon met de constatering dat de provincies niet vertegenwoordigd zouden moeten worden door een tussenpartij – zoals het IPO – maar liever zelf een belangrijke rol zouden spelen tijdens de evaluatie. Onderzoekers hebben provincies actief rollen toebedeeld om actieve betrokkenheid van de provincies en eigenaarschap over de evaluatie aan te moedigen. Wij onderscheiden het organiseren van (provinciaal) bestuurlijk commitment, het aanmoedigen van eigenaarschap en het bouwen van vertrouwen als de belangrijkste succesfactoren tijdens deze evaluatie periode. De onderzoekers hebben vanaf het begin relaties gelegd met provinciale bestuurders en bestuurlijke platforms om het belang van de lerende evaluatie extra kracht bij te zetten en ambassadeurs te identificeren die het belang van leren ondersteunden. Verder herkennen we het groeien van vertrouwen in de onderzoekers, middels transparant onderzoek en communicatie, en de sensitiviteit van de onderzoekers richting individuele provincies, als essentiële factoren. De aandacht die onderzoekers hebben besteed aan het bezoeken van individuele provincies, om helder te maken welke informatie nodig was en waarvoor, heeft in grote mate bijgedragen aan het vertrouwen. De onderzoekers zijn ervaren als toegankelijk en bereid om alle vragen te beantwoorden en mee te denken over zorgen. Ook het feit dat onderzoekers voorlopige

resultaten hebben gepresenteerd, en zich daarbij kwetsbaar opstelden, heeft bijgedragen en het opbouwen van de relaties tussen onderzoekers en de provincies.



Figuur 2. Overzicht van de factoren (zie ook linkerkant Figuur 1) die hebben bijgedragen aan het creëren van waarde ten aanzien van 1) stakeholder selectie, 2) het faciliteren van stakeholder betrokkenheid en 3) de afstemming tussen de beleidspraktijk en het evaluatieonderzoek.

De gevoelige bestuurlijke context en de zorgen om recentralisatie weerhieldd provincies dus in eerste instantie van actieve deelname aan de evaluatie. Interessant is dat deze angst voor afrekening er eveneens voor zorgde dat provincies aanwezig waren bij alle bijeenkomsten. Het voorzag de evaluatie van een zekere urgentie; een behoefte van provincies om een vinger aan de pols te houden bij de evaluatie. Enigszins paradoxaal heeft de verantwoordingsfunctie van evalueren op deze wijze bijgedragen aan de leerfunctie van evalueren. Wederom speelde de onafhankelijke status van het PBL hierbij een rol: meerdere malen hebben provincies aangegeven dat de status van PBL de bijeenkomsten van urgentie voorzag, wat hun motiveerde deze bij te wonen. Provincies geven te kennen dat zonder een partij als PBL, werksessies om ervaringen met betrekking tot natuurbeleid uit

te wisselen en te leren van elkaar niet tot stand zouden zijn gekomen, hoe waardevol deze ook worden gevonden.

Kortom, we zien dat de evaluatieonderzoekers met name actief gestuurd hebben op het realiseren van stakeholder betrokkenheid en continue afstemming tussen beleidspraktijk en het evaluatieonderzoek. Zeker stakeholder betrokkenheid vroeg veel aandacht en heeft in deze context meer invulling gekregen dan de noties van 'bereidheid' en 'urgentie' zoals uit het theoretisch kader. De onderzoekers hebben effectieve strategieën toegepast om betrokkenheid van de provincies te realiseren. Onverwacht was hoe de zorg voor recentralisatie en de status van PBL bijdroegen aan het realiseren van betrokkenheid. Dit is een waardevol inzicht dat gebruikt kan worden bij het vervolg van de evaluatie. Ook wat betreft continue afstemming tussen beleidspraktijk en evaluatieonderzoek heeft dit onderzoek bijgedragen aan een concretere invulling ten opzichte van het theoretisch kader. Eindgebruikers zijn actief betrokken geweest bij de opstelling van het evaluatiekader en er is actief gestuurd op regelmatige interactie en tussentijdse terugkoppeling van resultaten. Naast continue afstemming tussen beleidspraktijk en evaluatieonderzoek zijn er ook factoren te onderscheiden die hebben bijgedragen aan de ontwikkeling van een lerend beleidsnetwerk, zoals de interactiemomenten met diverse stakeholders en de manier waarop deze frame-reflectie en interactie aanmoedigden. Deze factoren vormen een aanvulling op het theoretisch kader.

5. Belangrijkste conclusies en aanbevelingen

Stakeholder selectie en betrokkenheid

Maatschappelijke partijen zijn beperkt betrokken geweest in de evaluatie tot dusver. Vanwege de toegenomen complexiteit van het natuurbeleid en het groeiend aantal actoren dat een rol speelt in beleidsvorming en -uitvoering is het echter van belang om te interacteren met andere partijen, omdat zij de kennis hebben die nodig is om het systeem beter te kunnen doorgronden. Wanneer deze partijen niet worden meegenomen, zal de ontwikkelde kennis minder sociaal robuust zijn, zal er minder draagvlak zijn voor de bevindingen en de toepassing ervan, wat uiteindelijk zal leiden tot verminderde beleidsimpact.

- Voor de volgende evaluatieperiode is het volgens ons raadzaam een bredere groep van stakeholders (zoals maatschappelijk partners betrokken bij de ontwikkeling en uitvoer van natuurbeleid, maar ook vertegenwoordigers van het Rijk) te betrekken bij het vaststellen van het evaluatiekader en het uitvoeren van het evaluatieonderzoek.

We hebben gezien dat er veel aandacht uit is gegaan naar het creëren van betrokkenheid van stakeholders bij de evaluatie. Een aantal factoren heeft daaraan bijgedragen zoals het identificeren van ambassadeurs en het creëren van vertrouwen door transparantie en de sensitiviteit van onderzoekers.

- Wanneer andere potentiële eindgebruikers een grotere rol krijgen bij de tweede evaluatie, moet er opnieuw aandacht besteed worden aan het realiseren van betrokkenheid van deze partijen en het creëren van vertrouwen in de evaluatoren. Zorgvuldige communicatie en transparantie, het identificeren van ambassadeurs, het organiseren van bestuurlijke steun zullen wederom belangrijke strategieën zijn. In het procesontwerp zal rekening gehouden

moeten worden met de toenemende complexiteit die gepaard gaat met een groter aantal actoren en meer heterogeniteit in perspectieven en belangen.

Afstemming tussen beleidspraktijk en evaluatieonderzoek

In alle fases van het evaluatieonderzoek zijn de eindgebruikers in meer of mindere mate betrokken geweest. In fase 1 (vaststellen evaluatiekader en methoden) is er veel interactie georganiseerd tussen onderzoekers en eindgebruikers, wat waardevolle inzichten heeft opgeleverd die voor een groot deel zijn meegenomen in het evaluatiekader, al is dit niet geheel systematisch gebeurd. Door regelmatige interactie tussen onderzoekers en beleidsbetrokkenen sluit de eindrapportage grotendeels aan bij de kennisbehoeften van de betrokkenen, met name als het gaat om kennis over planpotentieel, uitvoeringspotentieel, beleidsstrategieën en de agendering van de verbrede ambities. Volgens sommigen is door het gekozen schaalniveau (landelijk) het handelingsperspectief voor provinciaal niveau beperkt inzichtelijk. Dit hangt ook samen met de keuze voor de gebruikte onderzoeksmethoden (met name de Metanatuurplanner), die maar beperkt ter discussie stonden.

- Enerzijds bieden de eindrapportage en deze review al een set van uitgangsvragen voor het nieuwe evaluatiekader; het ligt voor de hand om de nulmetingen op te volgen met vervolgmetingen, na te gaan in hoeverre en op welke manier aanbevelingen vorm hebben gekregen in de praktijk (zie het evaluatie eindrapport), en aandacht te besteden aan het leren rondom de geïdentificeerde leerbehoeften (zie ook hieronder bij 'lerend beleidsnetwerk verder ontwikkelen'). Anderzijds is het voor de tweede evaluatie van het Natuurpact aan te bevelen fase 1 (vaststellen evaluatiekader en methoden) wederom zorgvuldig in te richten, waarbij de input van verschillende betrokkenen *systematisch* moet worden ingebed in het evaluatiekader. We bedoelen hiermee dat er voldoende tijd en ruimte beschikbaar dient te zijn in deze fase om vraagarticulatie door stakeholders te ondersteunen, hun input te analyseren en te vertalen naar gedeelde onderzoeksvragen.
- Vooral omdat het zal gaan om een groter aantal, wellicht tegenstrijdige, perspectieven en nieuwe kennis- en leerbehoeften bij andere partijen is het zaak dat het procesontwerp rekening houdt met deze belangrijke vertaalslag.
- Als gevolg van de verbreding van input door het betrekken van meerdere actoren, zullen ook de geschiktheid van gebruikte onderzoeksmethoden moeten worden bezien.

In fase 2 (data verzamelen) en fase 3 (data analyse / interpretatie) is er ook regelmatig interactie geweest tussen onderzoekers en beleidsbetrokkenen, wat heeft geleid tot onderling vertrouwen, met als gevolg een hoge kwaliteit van informatie, sociaal robuuste kennis en geaccepteerde bevindingen. De gekozen vormen van interactie hebben hier een belangrijke rol in gespeeld. Deze review heeft de waarde van verschillende vormen van interactie – bilaterale consultaties, groepsreviews, leersessies – voor verschillende doeleinden aangetoond.

- Met deze kennis kunnen dataverzameling en -analyse efficiënt en effectief worden ingevuld, op zo'n manier dat ze ook bijdragen aan het gezamenlijk leerproces. Daarnaast is het efficiënt en effectief gebleken om gebruik te maken van bestaande overlegstructuren, o.a. van provinciale beleidsmakers (zoals de Werkgroep Natuurbeleid).
- Regelmatige interactie blijft van belang in fase 2 en 3. Verschillende vormen van interacties (bijvoorbeeld bilaterale consultaties, groepsreview sessies) kunnen worden ingezet voor verschillende doeleinden, inclusief het gebruik van bestaande overlegstructuren.

De onderzoekers hadden de hoofdrol tijdens fase 4 (disseminatie); het eindrapport is nadrukkelijk een PBL rapport, en geen gezamenlijke productie. Dit werd door de beleidsbetrokkenen van groot belang geacht, omdat het de onafhankelijkheid van de onderzoekers, en daarmee de geloofwaardigheid van de resultaten, onderstreepten. Nadeel van deze strategie zou een gebrek aan draagvlak voor de bevindingen en aanbevelingen kunnen zijn. Door de regelmatige interacties in de eerdere fases was hier echter geen sprake van.

- De strategische waarde van een onafhankelijk PBL-rapport moet niet worden onderschat. De genomen strategieën om onafhankelijkheid te bewaken dienen te worden voortgezet.
- Tegelijkertijd kan onafhankelijkheid op gespannen voet staan met het benodigde draagvlak dat voortkomt uit een proces van co-creatie. Dit kan worden ondervangen door intensieve interactie in de eerdere fases van het evaluatieonderzoek.

Lerend beleidsnetwerk verder ontwikkelen

Een belangrijke bevinding is dat de lerende evaluatie van het Natuurpact netwerk waarde en affectieve waarde heeft opgeleverd; er heeft zich een lerende beleidsnetwerk gevormd waarin men uitdagingen, worstelingen, kansen en mogelijkheden met elkaar heeft gedeeld, met aan de basis daarvan onderling vertrouwen en wederzijdse geruststelling. De provincies en vertegenwoordigers van het Rijk hebben aangegeven het lerende karakter van de evaluatie te willen behouden. De combinatie van het belang van de evaluatie (wat zich vertaalde in de angst voor afrekenen) enerzijds en de mogelijkheid om te leren van andere actoren anderzijds, maakt dat de context van de evaluatie van het Natuurpact geschikt is (het geeft urgentie) om het leren rondom ontwikkelen en uitvoeren van natuurbeleid verder vorm te geven. Tijdens de eerste evaluatie is gebleken dat een aantal leerbehoeften onvoldoende zijn geadresseerd, wat consequenties heeft voor de gerealiseerde inhoudelijke en instrumentele waarde. Het gaat hierbij enerzijds om leerbehoeften van de primair betrokkenen (provincies), bijvoorbeeld met betrekking tot het concreter invulling geven aan de verbrede ambities, en bijbehorende indicatoren, en met betrekking tot de nieuwe rol als faciliterende overheid. Anderzijds gaat het om de leerbehoeften van andere actoren. Bijvoorbeeld, vertegenwoordigers van het Rijk hebben te kennen gegeven te willen leren over de invulling van hun nieuwe rol als systeemverantwoordelijke, zoals beschreven in het Natuurpact. Van andere actoren zijn de leerbehoeften nog niet in beeld.

- Het is raadzaam om de inhoud van de lerende evaluatie te laten meebewegen naar ontwikkelende leerbehoeften van de betrokken actoren, met aandacht voor onderwerpen waar nog veel meer te leren is en die tot nu toe buiten beeld zijn gebleven.
- Om daar zicht op te krijgen, alsmede op de leerbehoeften van 'nieuwe' actoren, kunnen de leervragen van deze partijen in beeld worden gebracht.
- Specifiek benadrukken we het in kaart brengen en adresseren van de leervragen van vertegenwoordigers van het ministerie van EZ (als (mede)opdrachtgever en tevens stakeholder)
- Het monitoren van de leeragenda's van betrokken actoren (en dus hun leervragen) vergroot het inzicht in de impact van de evaluatie op de beleidspraktijk. Bovendien biedt dit houvast voor invulling van de weer volgende evaluatieperiode (2021-2024). We raden aan het monitoren van leervragen expliciet op te nemen als een evaluatiedoel. Tussentijdse rapportages (halfjaarlijks) in de vorm van verschuivende leeragenda's zorgen voor tijdige feedback en reflectie.

- Om rekening te houden met nieuwe leervragen is er ruimte nodig in het procesontwerp van de nieuwe evaluatie; mogelijkheden tot het bijstellen van planning en budget als gevolg van nieuwe leervragen kunnen bijvoorbeeld opgenomen worden in de opdrachtovereenkomst.
- Een mogelijke keuze voor een meer integrale insteek van de evaluatie, gericht op het geheel van Nederlands natuurbeleid in plaats van uitsluitend de afspraken uit het Natuurpact, brengt ook nieuwe kennisbehoeften en een vraag naar aanvullende expertises met zich mee. We raden aan hier bewust van te zijn.

Balans blijven zoeken tussen leren en verantwoorden

De resultaten tonen aan dat er een fragiele balans is gevonden tussen leren en verantwoorden. Diverse leerprocessen hebben zich voltooid bij de primair betrokkenen van de evaluatie. Tegelijkertijd zien we dat de evaluatie succesvol wordt gebruikt voor verantwoording van natuurbeleid op diverse overheidsniveaus. We zien dat leren en verantwoorden elkaar zowel versterken als tegenwerken. De gespannen verhoudingen tussen de provincies en het Rijk, en het feit dat sommige provincies PBL zien als een verlengstuk van het Rijk, maakten dat provincies verantwoorden vertaalden naar een zorg om afrekenen. De evaluatie kreeg hierdoor urgentie, wat bijdroeg aan de hoge opkomst van provincies tijdens evaluatie bijeenkomsten, met een positief effect op hun leerprocessen als gevolg. Anderzijds heeft de angst voor afrekenen ervoor gezorgd dat betrokkenen niet altijd het achterste van hun tong lieten zien; ze waren soms terughoudend in het delen van informatie, wat het evaluatie proces bemoeilijkte. De gespannen verhoudingen tussen de provincies en het Rijk lijkt gedurende de evaluatieperiode te zijn verbeterd. Ook het PBL heeft aan vertrouwen gewonnen deze evaluatieronde. Tegelijkertijd zou het naïef zijn te denken dat de angst voor afrekenen verdwenen is. Zeker gezien de aard van de tweede evaluatie, waarbij er niet alleen *ex-ante* wordt getoetst, maar ook *ex-post*, zal de verantwoordings-toets eerder een grotere dan een kleinere rol gaan innemen. Door het *ex-post* karakter zal meer worden gekeken naar de werkelijke resultaten die de provincies hebben geboekt, wat potentieel de zorg om afrekenen en recentralisatie hoger doen oplopen. Dit heeft onvermijdelijk consequenties voor hoe leren en verantwoorden zich tot elkaar verhouden, en daardoor ook voor de opgebouwde relaties tussen onderzoekers en eindgebruikers van de evaluatie.

- Om de balans te bewaken tussen leren en verantwoorden is het raadzaam om deze beide concepten op te nemen in het procesontwerp van de evaluatie. Een voorbeeld hiervan is de scheiding tussen teamleden die zich primair richten op het verantwoorden van beleid (en daarmee wetenschappelijke kwaliteit en onafhankelijkheid) en teamleden die zich primair richten op interactie met beleidsbetrokkenen, zoals werd aangehouden in de afgelopen evaluatie periode.

Tot dusver heeft verantwoorden met name betrekking gehad op de biodiversiteitsambitie van het natuurbeleid. De beschikbare beleidskaders geven hiervoor urgentie en sturing. Als gevolg hiervan lag de nadruk op het realiseren van hectares en mogelijke perverse prikkels. Voor de twee nieuwe ambities staan er nog geen kaders vast en is er dus de mogelijkheid deze meer gezamenlijk invulling te geven. Er is behoefte aan concrete, niet-vaststaande doelen en relevante, niet-rigide indicatoren; in andere woorden, er is behoefte aan doelen en indicatoren die houvast bieden voor verantwoording, maar voldoende ruimte over laten om te leren en bij te sturen.

- Het is raadzaam deze doelen en indicatoren gezamenlijk te ontwikkelen met relevante stakeholders (zowel provincies als maatschappelijke partners). Zonder hun betrokkenheid is er een risico op een gebrek aan eigenaarschap van de doelen en de indicatoren, en daarmee een vergrote kans op perverse prikkels.

Er is vooral sprake geweest van verantwoording richting hogere overheidsniveaus (opwaartse verantwoording), in mindere mate richting maatschappelijke partijen. Een meer horizontale vorm van verantwoording zou inhouden dat provincies natuurbeleid ontwikkelen dat rekening houdt met de perspectieven en belangen van maatschappelijke partijen. Zoals we al eerder schreven is het een risico maatschappelijk partners beperkt te betrekken tijdens de evaluatie. Zij zijn niet meegenomen in de leerprocessen en de daarmee gepaarde ontwikkeling van het gedachtegoed over de verbrede ambities van natuurbeleid. Het achterblijven van de leerprocessen van maatschappelijke partijen heeft consequenties voor het realiseren van horizontale verantwoording; de perspectieven en belangen van deze partijen zijn beperkt inzichtelijk en divergeren mogelijk van het beoogde beleid. Daarnaast zijn provincies in belangrijke mate afhankelijk van maatschappelijke partners voor het ontwikkelen en uitvoeren van natuurbeleid en dus het realiseren van de ambities; horizontale verantwoording kan daardoor een positief effect hebben op doelbereik en op opwaartse verantwoording, mits er in voldoende mate sprake is van een gedeelde ambitie.

- Zoals we al eerder benadrukten is het van belang maatschappelijk partners mee te nemen in de volgende evaluatie periode en de daarmee gepaarde ontwikkeling van een verbreed perspectief, ten goede van sociaal robuuste kennis en draagvlak, maar ook voor de balans tussen leren en verantwoording. Zoals we al eerder stelden, is het raadzaam om in het procesontwerp rekening te houden met de toenemende complexiteit die gepaard gaat met een groter aantal actoren en meer heterogeniteit in perspectieven en belangen. Naast organisatorische complexiteit en het omgaan met grote hoeveelheden, wellicht tegenstrijdige, informatie en perspectieven, gaat het ook om de bereidheid om kennis te delen én de veiligheid om dat te kunnen doen.
- Ten bate van deze veiligheid is het zinvol af te wisselen tussen homo,- en heterogene groepen. Deelnemers voelen zich doorgaans veiliger in homogene settings. Door daar mee te starten kunnen nieuwe partijen in een veilige omgeving bekend(er) worden met de evaluatie en een band opbouwen met de onderzoekers, alvorens te interacteren met andere stakeholders. Diverse consensus methodieken kunnen behulpzaam zijn (e.g. Delphi-studies, dialogosessies, etc.), hoewel niet alle verschillen in perspectieven hoeven worden opgelost.

De lerende evaluatie van het Natuurpact heeft veelbelovende uitkomsten laten zien. Gezamenlijke leerprocessen, gedeeld tussen de provincies, maar ook het Rijk en maatschappelijke partners, zijn in gang gezet. De komende jaren kunnen deze processen worden voortgezet, waarbij sociaal robuuste kennis wordt ontwikkeld die de beleidspraktijk verder kan verrijken. Daarnaast kan de lerende evaluatie ook in de toekomst bijdragen aan het verder vormen en verdiepen van het lerend beleidsnetwerk. Dit biedt perspectief voor het vergroten van de beleidsimpact van natuurbeleid om zo de resterende opgaven en de verbrede ambities te realiseren in 2027.

1. Introduction

1.1 Origin of the Natuurpact agreement and its evaluation

Dutch nature policy has been decentralised, making the 12 provinces responsible for both its development and its implementation. Meanwhile, national government has remained accountable to the European Commission (EC) for realising internationally agreed upon nature goals, and is responsible for providing national government policy frameworks. The agreements on the decentralisation and the ambitions of Dutch nature policy are set out in the Coalition Agreement Nature (2011/2012) and the Natuurpact (2013). Provincial and national governments have agreed to collaborate to complete the Dutch Nature Network, achieve the international nature goals, increase societal engagement with nature and promote the relation between nature and the Dutch economy. These ambitions are to be realised by 2027.

As a consequence of the recent developments (not only decentralisation, but also horizontalisation in the form of Europeanisation), Dutch nature policy is increasingly characterised by multi-stakeholder involvement and multi-level governance, contributing to its inherent complexity. Furthermore, the high ambitions outlined in the Natuurpact demand an integrated and collaborative approach to policymaking. Hence, national government and the provinces together decided that the evaluation of the Natuurpact ambitions should allow for learning-by-doing and informing policy processes *during* policy development and implementation on multiple governmental levels. The Ministry of Economic Affairs (Dutch: Ministerie van Economische Zaken, EZ) and the Association of the Provinces of the Netherlands (Dutch: Interprovinciaal Overleg, IPO) have commissioned the Netherlands Environmental Assessment Agency (Dutch: Planbureau voor de Leefomgeving, PBL) to conduct a learning-focused evaluation. The PBL has partnered with Wageningen University and Research (WUR) to conduct this assignment.

It was decided to adopt a reflexive evaluation approach, which has a strong focus on participation, occurs *during* the policy process, and combines the learning and accountability purposes of evaluation. By applying this approach, PBL is to report on the progress of realising the Natuurpact ambitions every three years: the first report was published in January 2017³.

This introduction further explains the call for a reflexive evaluation approach in the context of Dutch nature policy, followed by a theoretical explanation of reflexive



Front-page of the final evaluation report (PBL & WUR, 2017)

³ All reports, including the background reports, can be downloaded via: <http://themasites.pbl.nl/evaluatie-natuurpact/>

evaluation. Then, we highlight the purpose and focus of this review of the Natuurpact reflexive evaluation, including our main research questions.

1.2 Call for reflexive evaluation

1.2.1 Developments in Dutch nature policy

As touched upon in the previous paragraph, two major developments in the context of nature policy gave rise for the call for reflexive evaluation. First, the political and administrative context has changed in recent years. Authority and responsibility for nature policy have shifted from national government to the provinces (decentralisation) and towards the European Union (EU) (Europeanisation). Nature policy is thus increasingly characterised by multi-level governance, and requires national government and the provinces to give substance to their newly acquired roles and responsibilities. National government has retained its responsibility to the EU for achieving internationally agreed upon nature goals and for reporting on national progress with regard to these goals, while the provinces have become responsible for the development and implementation of nature policy. Furthermore, nature policy has increasingly become a multi-actor matter. To an ever greater extent, policies are developed and implemented in collaboration with more local societal parties. Second, the ambitions of Dutch nature policy regarding biodiversity are of a significant character, to which two major ambitions have been added: increasing societal engagement with nature and strengthening the ties between nature and the Dutch economy. For these new ambitions, policies need to be further tested in order for these to develop.

The PBL and the commissioners reasoned that together these developments called for an approach to evaluation that promotes collaboration and opportunities to learn from experiences of implementing different policies, while also allowing for gaining insight in the effectiveness of implemented policy strategies. They expected such an approach would prevent each province to individually re-invent the wheel, and simultaneously would provide insights that could directly be used to timely adapt policy plans and their execution to increase the likelihood of achieving the ambitions by 2027. Hence, the PBL and the commissioners agreed to employ a reflexive evaluation approach.

1.2.2 Reflexive evaluation

Scholars argue that participative research is better aligned to societal needs, and produces enriched knowledge that is societally robust because it is recognisable, perceived as (scientifically) reliable and applicable by society (Nowotny, 2000; Lang et al., 2012). From a policy evaluation perspective, by employing a utilisation-focused approach to evaluation and involving policymakers and other relevant stakeholders in determining its design and scope, policy practice becomes knowledge-enriched, thereby eventually leading to greater policy impact (Edelenbos & van Buuren, 2005; Patton, 2000). Impact is further increased as reflexive evaluation takes place *during* the policy process: it allows for timely adjustments and so increases the likelihood of attaining the relevant goals.

With this utilisation-focused approach and its participative and collaborative character, reflexive evaluation aims to take into account complexities arising from multi-actor and multi-level governance by seeking to unite two important functions of evaluation: accountability and learning. Though it may appear that these the two purposes run counter to each other (e.g. national government demands a quick impact assessment for accountability to the EC, while provincial governments require space for learning and experimenting), in reflexive evaluation there is attention to both (e.g. impact assessment

is conducted to optimally inform provincial learning processes, allowing adaptation and learning, while also suiting the national government's accountability to the EC). Conceptually, evaluation for accountability and learning may be understood as two sides of the same coin, overcoming the proposed dichotomy (Guijt, 2010; Regeer, De Wildt-Liesveld, van Mierlo, & Bunders, 2016).

Moreover, the new ambitions drawn up in the Natuurpact agreement may be viewed as an ambition for system innovation towards sustainability. Holding this system-thinking perspective, system learning (e.g. identifying barriers and creating opportunities to overcome these to establish system change) and enhanced reflexivity as valuable resulting asset, are perceived as paramount for promoting transformation to achieve sustainability (Elzen, Augustyn, Barbier, & Mierlo, 2017). Reflexive evaluation is an approach that seeks to achieve reflexivity and system learning (Arkesteijn, van Mierlo, & Leeuwis, 2015).

1.3 Review of the Natuurpact reflexive evaluation

1.3.1 Purpose of this review

For PBL and partner WUR, conducting such a large-scale reflexive evaluation with a high level of participation is relatively new. To ensure scientific rigour of this methodological innovation, PBL commissioned the Athena Institute (VU University Amsterdam, the authors of this review) to construct a theoretical framework comprising the key elements of reflexive evaluation (Van Veen, Verwoerd, & Regeer, 2016)⁴, based on an extensive literature study. This framework provided the basis for the evaluation researchers to steer the evaluation process, with the support of the present authors. This review presents how this has yielded expected and unanticipated types of value for the participants of the evaluation. Here, we draw lessons from the successes and shortcomings in order to refine the theoretical framework for the continuation of the evaluation towards 2027.

1.3.2 Demarcations: focus on process

The reflexive evaluation aims to contribute to policy processes and thus achieve the impact of policy on society to address complex societal issues. The scholarly literature has discussed the difficulties of evaluating studies that strive to solve complex social problems (Jahn & Keil, 2015): how can the quality of such research be determined? How do we define policy impact? How can it be made measurable, how do we increase its scale? How can societal developments be unambiguously related to policy changes originating in the reflexive evaluation? The latter question is especially complex due to the timing of review: too soon, and policy impact is likely to be limited, too late and demonstrating causality becomes impossible (Walter, Helgenberger, Wiek, & Scholz, 2007). It is due to these difficulties that scholars recommend focusing on *the process* of the research, rather than on its (eventual) outcomes.

In view of these complexities in quantifying quality and the recognition our study took place too soon after the publication of the final evaluation reports to fully map all the effects on policy practice, it was decided that this report should focus on the *process* of the Natuurpact reflexive evaluation over

⁴ This literature study may also be downloaded via: <http://themasites.pbl.nl/evaluatie-natuurpact/wp-content/uploads/characteristics-of-reflexive-evaluation.pdf>

the last four years. Nevertheless, despite the slightly premature nature of our evaluation, we take into account a number of outcomes the evaluation researchers anticipate.

The PBL anticipated that reflexive evaluation would produce high quality knowledge that is scientifically sound and societally robust. Though we draw no conclusions on the scientific soundness of the knowledge that was generated, we do reflect on its social robustness. Also, the evaluation researchers expected the *ex-durante* character would improve the timeliness of the evaluation findings, making it possible for policymakers to adapt policy along the way towards 2027, thus increasing the likelihood of realising the Natuurpact ambitions. In this report, we show that indeed some provinces have adapted their plans as a result of the Natuurpact evaluation.

1.3.3 Main research questions

The objective of this report is to formulate recommendations to PBL for the continuation of the reflexive evaluation (I) by providing insight into how (participating with) the evaluation has been of value to the policy practice of national and provincial policymakers and how this value relates to the expected outcomes, and (II) by determining which factors promoted or inhibited value and how evaluation researchers have acted upon these factors.

This objective translates into two main research questions:

- I. In which ways has the Natuurpact reflexive evaluation been of value for national and provincial government policy practice and how does this relate to the anticipated outcomes of knowledge-enriched policy practice and multi-directional accountability?
- II. How has steering based on the constructed theoretical framework (Van Veen et al. 2016) contributed to establishing these values and which other unexpected factors may be distinguished that influenced the evaluation process?

1.3.4 Research approach

It is important to note that to make the study feasible, we focused on the experiences of the primary end-users of the evaluation: national and provincial policymakers. Although societal partners and businesses have been involved during the evaluation, their role in shaping its design and scope was relatively small. In chapter 6.1 we further discuss stakeholder involvement and diversity. Our data was collected from rounds of in-depth interviews with national and provincial policymakers, and a focus group with provincial policymakers. Furthermore, we have analysed our observations during interactions between the participants and the evaluation researchers over the last four years.

Our focus on the end-users of the evaluation also implies that the experiences of the evaluation researchers – how the evaluation has been of value to their research practice, the challenges they faced and how they dealt with these – are discussed only indirectly. In a following evaluation study, there will be focused on the value of reflexive evaluation to knowledge production and lessons learned from the Natuurpact evaluation for evaluation researchers.

Another aspect that deserves consideration is the dual role of the current authors. The Athena Institute has been intensively involved in shaping the evaluation's design: its process, the organised

interactions with the participants such as the workshops and group-review sessions. In this report, we reflect on this process to determine whether it was successful in establishing the intended value – rather as if we are marking our own paper. To guarantee our independence and an unbiased review, we collaborated in our analysis with colleagues from our institute who have no further involvement in the Natuurpact evaluation. Furthermore, external experts in the fields of policy and evaluation reviewed our theoretical framework (Van Veen et al., 2016) and this report to ensure its scientific soundness.

1.4 Reader's guide

In the following chapters, we start out with a description of the Natuurpact evaluation as it occurred; the Natuurpact ambitions, the evaluation's main objectives and research questions, the actors involved and the main events that took place in the evaluation research phase. In chapter 3 we present the theoretical framework (after Van Veen et al. 2016) and its operationalisation for the purpose of this study. Chapter 4 comprises our methodology, and in chapters 5 and 6 we present the results of the analysis; the value of reflexive evaluation and factors that led to these values, respectively. In chapter 7 we discuss our findings in relation to the main research questions, draw main conclusions and present recommendations for the continuation of the Natuurpact reflexive evaluation.



Characteristic Dutch landscape

2. The Natuurpact evaluation

This chapter provides additional background on the context in which the reflexive evaluation was conducted, describes the actors who were involved and how the evaluation (in terms of e.g. sub-projects and related workshops) was conducted.

2.1 Setting the scene: the context in which the evaluation took place

The two major developments (i.e. increase multi-level governance and multi-actor character, and the broadened ambitions) in nature policy context that gave rise to the call for reflexive evaluation have already been touched upon in the introduction. In this chapter we provide additional context, which had significant influence on how the different parties behaved and participated in the evaluation.

2.1.1 History of relations between national and provincial governments

Without diving too far back into the history of Dutch nature policy, the relation between national government and the provinces deserves consideration as it was – and is – an important determinant of the evaluation process. Prior to decentralisation and the Natuurpact agreement, the provinces were responsible for implementing nature policy and were accountable to national government. These lines of accountability were strict. In the agreements on the decentralisation (of which the Natuurpact was the final accord, and the result of a long process of negotiation between the governments) it was agreed that the vertical lines of accountability would cease: the provinces are autonomous and from then on would be accountable only to the Provincial Executives. However, the Natuurpact also underlines that national government remains system responsible. Furthermore, national government is accountable to the EC for reaching the biodiversity targets recorded in the VHR and the KRW. National government thus wished to remain informed on provincial progress on the biodiversity goals and therefore, together with the IPO, negotiated the three-yearly evaluation. The decentralisation and its effect on nature policy (and nature) were explicitly not an evaluation topic. EZ (as representative of national government) and IPO commissioned PBL for this task, specifically asking for a participative type of evaluation that would allow for learning. In conducting this evaluation, the PBL partnered with the WUR.

Important to note here is that, though IPO is in principle a representative of the provinces, it is not recognised as such by the provinces themselves. IPO sat at the negotiation table on decentralisation as representative of the provinces, but was recalled several times for insufficiently guarding the provinces' interests. The provinces perceive IPO to be too close to EZ and approach it with the same levels of distrust. In other words: EZ and IPO commissioned a participative evaluation approach in a top-down fashion. To the provinces, which were expected to participate, this felt enforced. The involvement of EZ with the evaluation further troubled the provinces as they interpreted it as EZ dishonouring the agreement on ceasing upwards accountability lines. Moreover, the provinces were concerned that EZ would use the evaluation to eventually legitimise the re-centralisation of nature policy. That PBL was commissioned for the evaluation also did not help in this regard – some provinces perceived PBL (an independent organisation in practice, but formally a sub-department of the Ministry of Infrastructure and Environment) as an extension of national government.

2.1.2 Top-down commissioned participative evaluation

Although mostly implicit and rarely to the fore, the top-down evaluation assignment and the distrust between the formal commissioners and the provinces (and by extension, towards PBL) had major implications for the levels of participation of some of the provinces with the reflexive evaluation. While no provinces openly refused to participate, most were at times apprehensive and some even opposed to sharing detailed information on their policy plans, concerned about the potential retributions (additional cutbacks, re-centralisation). It resulted in strategic play – withholding information, either painting a more negative or positive picture of the situation in their province – to send a signal to their administration and national government. Although this played a larger role for some provinces than for others, in general it restricted the extent with which they were willing to actively and openly participate in the evaluation. We discuss in chapter 6 how the evaluation researchers anticipated this, and the importance of their actions for conducting the reflexive evaluation.

2.2 Actors involved with the evaluation

The primary actors involved with the Natuurpact evaluation were the evaluation participants and the evaluation researchers. Figure 2.1 shows the decision-making bodies involved in the evaluation across the different levels of government and which bodies interacted most directly with the evaluation researchers.

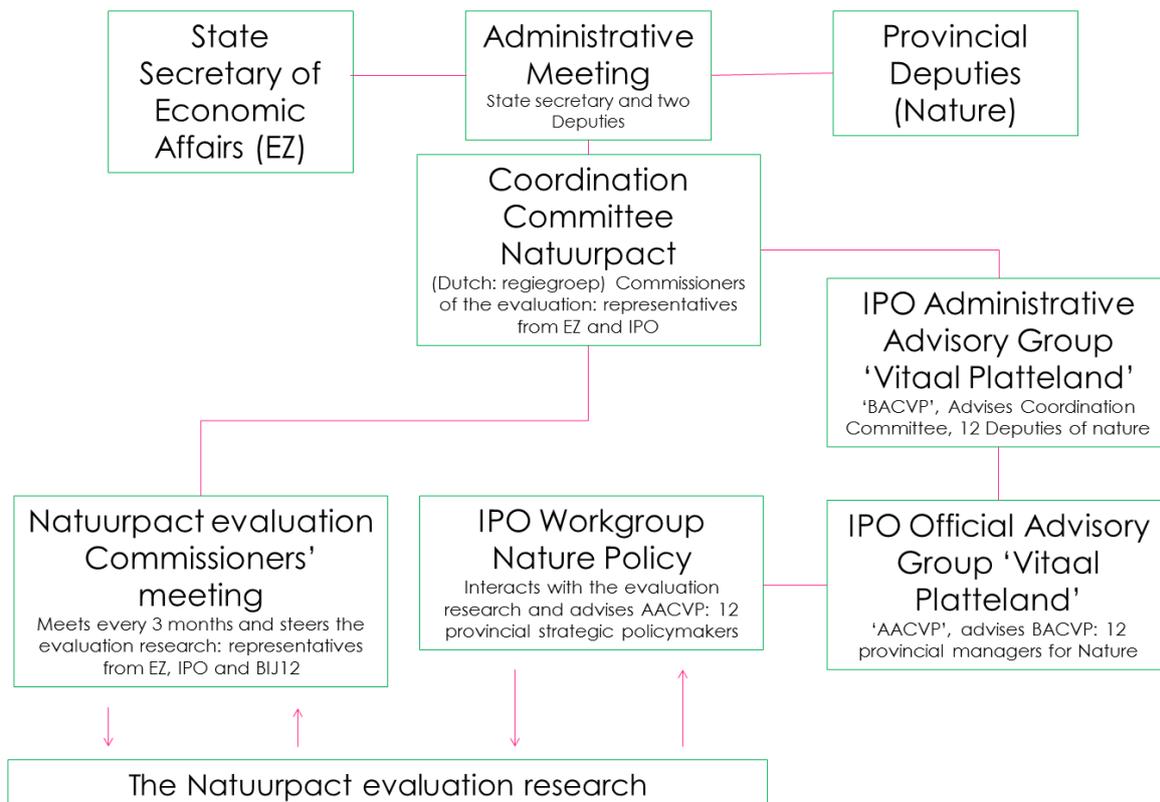


Figure 2.1 The decision-making bodies of the Natuurpact evaluation across different levels of government. The evaluation researchers interacted frequently with the Commissioners' meeting (every three months) and with the Workgroup Nature Policy (approx. once every 1-2 months, more frequently with two appointed representatives from the Workgroup).

2.2.1 The evaluation participants

The formal commissioners: EZ and IPO

The Natuurpact evaluation was formally commissioned by EZ and the IPO. Representatives from these parties are in the Coordination Committee and the Commissioners' meeting (Figure 2.1), through which they ensured the evaluation research fulfils the original assignment. These representatives attended the workshops, the group-review sessions, and other administrative meetings in which the evaluation was discussed.

The IPO formally represents and safeguards the collective interests of the provinces to The Hague and Brussels. It does so through informing and supporting provincial policy processes and by facilitating knowledge exchange between the provinces and societal partners and other stakeholders (supported by BIJ12, the executive wing of IPO also involved in provincial data management). As discussed previously, the provinces have only limited acceptance of IPO as their formal representative, and regard the organisation with a level of distrust. The IPO offers a platform for interprovincial knowledge exchange to contribute to the quality, effectiveness and efficiency of the provincial administration (IPO, 2017). As such, the IPO (together with BIJ12) had a supporting role in the execution of the Natuurpact in bringing actors into contact with each other and providing data for the evaluation.

The provinces

The largest group of end-users were provincial policymakers from the nature sectors. During the course of the evaluation, they were overall represented by the Workgroup Nature Policy (set up by the IPO, Figure 2.1). This workgroup comprises a strategic policymaker from each province, as well as a chairperson (a representative from province Gelderland) and a secretary from the IPO. The Workgroup deals with topics in nature policy that concern multiple provinces simultaneously, at a strategic level, including the Natuurpact evaluation.



Cakes served during collective workshops, illustrating the multiple government levels that play a role in Dutch nature policy

Through frequent meetings with the Workgroup, the evaluation researchers coordinated with provincial policy practice. From the Workgroup, two policymakers were chosen to take on a more active role in ensuring this alignment. They met with the evaluation project team more frequently to further the evaluation research.

Other provincial policymakers (with either more operational or strategic tasks) were also involved during the evaluation. They attended the workshops and participated in other research-related activities. Who and how many representatives attended for each province was always left to the provinces themselves. Sometimes there was only one representative of a province, at others this encouraged policymakers from the water sector to join. Provincial policymakers were the main participants in most of the workshops and review sessions and therefore had a larger input to the evaluation scope than did other stakeholders

Societal partners

Societal partners play an important role in provincial nature policy. Such partners include, for instance, terrain management organisations (e.g. Natuurmonumenten and Staatsbosbeheer), water boards, agricultural representatives (e.g. LTO) and non-government organisations (NGOs) (e.g. Vogelbescherming), to name but a few. Their role in determining the evaluation's scope was smaller than that of the commissioners and the provinces. They were primarily involved in data collection and for case studies, and participated in workshops during which evaluation findings and their implications were discussed.

2.2.2 The evaluation research institutes

PBL Netherlands Environmental Assessment Agency

PBL is an independent national institute for policy evaluation, specifically for domains related to environmental and nature policy, and is organisationally part of the Ministry of Infrastructure and Environment. PBL has the legal status of a policy analysis agency and is a prominent authority in the relevant fields. Historically, PBL is a more distant and traditional policy evaluator, but in view of the developments in contemporary public policy (e.g. increasingly multi-actor and multi-level governance) recognises the need for methodological innovation and is therefore exploring new methods, such as reflexive evaluation. From the PBL, researchers from different disciplinary backgrounds, such as ecology and public administration, were involved in the reflexive evaluation. Some of these were specifically recruited for their skills and experiences with participative research, to ensure the project team comprised sufficient expertise to undertake the evaluation.

Wageningen University and Research (WUR)

The WUR is a familiar research partner of PBL in environmental policy studies. For this evaluation, WUR researchers with expertise in participative and learning-oriented evaluation approaches were assigned, which proved an important source of knowledge to inform the design of the Natuurpact evaluation. The WUR conducted a number of components of the evaluation, and also had a significant role in shaping the overall evaluation approach.

The Athena Institute (VU University Amsterdam)

The authors work at the Athena Institute from the VU University Amsterdam, a research institute specialised in studying the interactions in the interface of science, technology and society, and participative research. As touched upon in the introduction, we played a somewhat dual role during the evaluation. We were commissioned by PBL to advise on designing the reflexive evaluation and shaping the interactions between the evaluation researchers and the participants. In addition, PBL asked us to scientifically review their approach to reflexive evaluation, of which the current report is an output. To conduct this assignment, the VU research team undertook a detailed literature study to build a theoretical framework on reflexive evaluation (Van Veen et al., 2016), on which this report is based.

We are aware of our potential (unintended) bias in reviewing a process to which we actively gave shape and have taken several steps to address this. These, for instance, included triangulation (of data

and researchers) and external reviews of our reports by experts in the fields of (environmental) policy and participative evaluation. We discuss this further in our methods chapter.

2.3 The Natuurpact evaluation scope

2.3.1 The formal evaluation assignment

EZ and IPO commissioned PBL to assess the effectiveness and efficiency of provincial nature policy, and how this is affected by prevailing government frames (e.g. laws and regulations). Furthermore, they asked the researchers to propose action perspectives and policy options that would enable the nature policy ambitions to be achieved more effectively and efficiently. The effectiveness and efficiency of decentralisation was explicitly not a topic for study during this evaluation. Finally, in the formal evaluation assignment it was decided a learning-oriented evaluation approach would be adopted.

PBL and the WUR will evaluate nature policy every three years. The first evaluation (on which this current review focuses) centres around nature policy and how this functions in relation to the legislative and policy frames originating from provincial, national and European government, as well as the potential contribution of nature policy to achieving the three ambitions formulated in the Natuurpact.

The evaluation assesses the *potential* of policy strategies (*ex-ante* evaluation, before implementation). This was decided as the provinces are at the start of a new implementation period – it was believed to be too soon to fully assess actual policy impact. Nevertheless, the evaluation does include the first experiences with policy implementation so far (from decentralisation to date), to gain insight into the feasibility of policy plans and the experiences with interactions with the government frames. The (cost)efficiency of nature policy was not included during this first evaluation because national and provincial governments decided it was too soon to study this. Moreover, the available financial data was of insufficient detail and differed too much between provinces to be used in a single assessment.

The evaluation assesses the joint (potential) effects of the provincial policy strategies in relation to the ambitions. In the main evaluation report these national effects are discussed (PBL & WUR, 2017). It is important to note that the ambitions are formulated at the national level – these are not translatable to provincial scale. As a result, the evaluation provides limited action perspectives or policy options at the provincial level. Nevertheless, a series of background reports discuss findings in more detail, including at the provincial level.

2.3.2 The three Natuurpact ambitions

Prior to the start of the evaluation – during phase I, which we further discuss later – the evaluation framework was jointly determined (PBL & WUR, 2015). National government, the provinces and societal partners agreed that the Natuurpact evaluation should assess the potential of nature policy to achieve three major ambitions:

- Increase biodiversity
- Increase societal engagement with nature
- Strengthen the relation between nature and the Dutch economy

Improving biodiversity relates to the conservation and increase of variety in organisms and their ecosystems. Specifically, the goals formulated in the European Bird and Habitat Directives (in Dutch, Vogel- en Habitat Richtlijnen, VHR) and the Waterboard Directive (in Dutch, Kaderrichtlijn Water, KRW) are assumed to encompass this ambition. Increasing social engagement with nature is an ambition that has not been articulated in specific targets. The central focus is the desire of provinces and national government to anchor nature in society, create a more solid foundation for nature and to share responsibility for nature with society. Strengthening the relation between nature and the economy relates to increasing the importance of nature for economic development, and also increasing the contribution of businesses to nature. Specific targets against which to measure policy have not been established regarding this ambition.

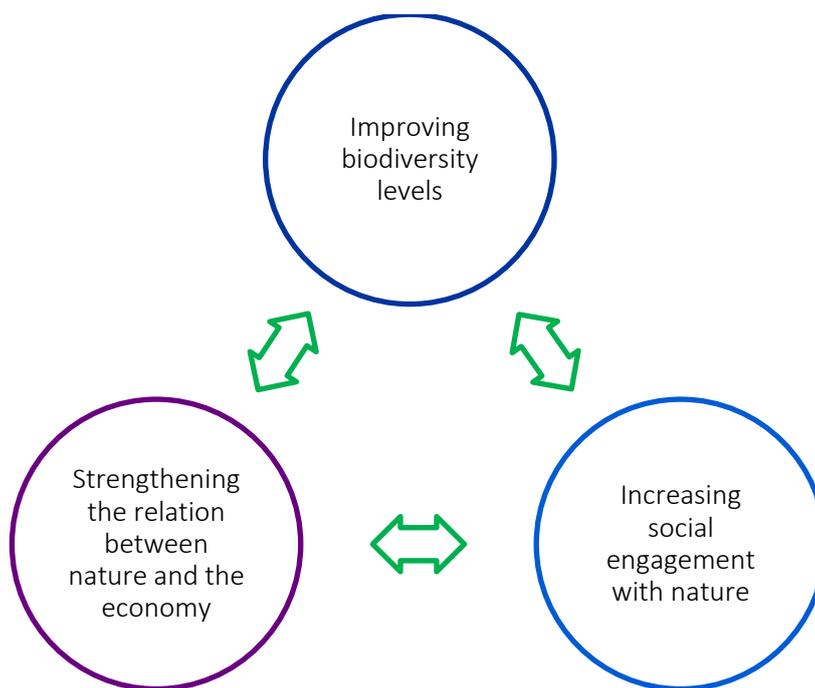


Figure 2.2 The three Natuurpact ambitions

The ambitions are posed as stand-alone but strongly interrelated – there may be synergy between them, but there may also be trade-offs. This evaluation did not address interaction between the ambitions, as there are no specific targets against which to assess policy regarding social engagement and the relation between nature and the economy.

2.4 The evaluation design

The reflexive evaluation aspired to emphasise learning from evaluation. Theoretically, it is a participative evaluation approach during which researchers and participants jointly conduct all research phases to co-create knowledge that is scientifically sound and optimally relevant to inform policy practice. To this end, researchers and the participants of the evaluated policy assess whether it is sufficient to attain goals and collectively develop action perspectives and policy options. Furthermore, the evaluation is developmental, meaning that the design allowed for development and

adaptation to issues emerging from policy practice to optimally adhere to the participants' needs for information.

In this section, we describe the evaluation research as it occurred, structured according to four major phases: determining the evaluation framework, data collection and interpretation, data analysis and interpretation, and the formulation of action perspectives and policy options (and dissemination). The rationale behind certain decisions are discussed in results sections of this report – this description serves as contextual background to better interpret the findings later on.

We emphasise that to fully describe how the evaluation progressed goes beyond the purpose of this review. We are aware that our description of how the evaluation took place is not exhaustive and does limited justice to all the research that occurred, e.g. all moments of interaction and discussion, important decisions made. For sake of coherence, we focus on describing the evaluation as it occurred and the events that are important to reflect on crucial factors that promoted or inhibited its success and how the researchers anticipated these.

2.4.1 Phase 1 – Determining the evaluation framework

Preliminary research

Although the Natuurpact discusses ambitions and goals of nature policy, at the outset of the evaluation it was undecided exactly which questions it should answer and what the scope of the research should be. For this reason, PBL and the WUR conducted a preliminary study. This included an inventory of provincial policy (Kuindersma et al., 2015)⁵, interviews with provincial policymakers and Deputies of the nature sectors, and with representatives from societal partners such as national terrain managing organisations (e.g. Natuurmonumenten). These interviews were used to gain a first insight into what the intended participants of the evaluation perceived as its purpose, preferred scope and demarcations.

3 'Learning Sessions'

Building on this preliminary work, three Learning Sessions were organised. During these sessions, the participants discussed the ambitions and the corresponding nature policy theory for biodiversity and, to a lesser extent, the other two ambitions. Based on these theories it was further discussed which policy strategies should be the main focus of the evaluation, and against which nature policy goals these strategies should be assessed. In addition, the government frameworks and how these affect policy strategies were selected. Table 2.1 shows the three Learning Sessions, the purpose of each and who attended (in the first session only policymakers were invited).

⁵ Kuindersma, W., F.G. Boonstra, R.A. Arnouts, R. Folkert, R.J. Fontein, A. van Hinsberg & D.A. Kamphorst (2015). *Vernieuwing in het provinciaal natuurbeleid; Vooronderzoek voor de evaluatie van het Natuurpact*. Wettelijke Onderzoekstaken Natuur & Milieu, Wageningen UR. WOt-technical report 35. 74 p.; 6 tab.; 23 ref.

Table 2.1 Overview of three Learning Sessions and their main characteristics

Learning session	Objective	Parallel sessions	Attendees
I (9/10/2014)	Prioritising nature goals and drafting a policy theory	Dutch Nature Network; Nature, Society and Economy; Agricultural nature management and species conservation; Conditions, environment and water	Commissioners EZ and IPO Provincial policymakers
II (1/8/2015)	Validation of policy theory and inventorying evaluation research questions and strategies	Biodiversity; Nature, Society and Economy	Commissioners EZ and IPO Provincial policymakers Societal partners
III (2/12/2015)	Collecting information needs and indicators suitable to conduct the evaluation	European Birds and Habitat Directives; European Water Framework Directive; Nature Quality, Nature Network and Biodiversity; Case studies for new policy strategies	Commissioners EZ and IPO Provincial policymakers Societal partners

The evaluation framework was then developed on the basis of the preliminary work and the three Learning Sessions (PBL & WUR, 2015). It set out the goals against which the selected policy strategies are assessed, the evaluation research questions, and the selected government frames and their interaction with the selected strategies. It also included an outline of the research approach in terms of its methodology and planning. The evaluation plan was considered to be a ‘living’ document, meaning it was anticipated that the plans would be further developed in response to the needs for information in policy practice.

There were two types of policy strategy selected for assessment. First, the ‘regular’ strategies, which are assessed for their contribution to biodiversity levels. The second category concerns ‘innovative policy strategies’, which are both existing strategies for biodiversity and novel strategies for engaging society and the economy, and were jointly selected by the researchers and participants. These are explored for their potential to contribute to biodiversity, societal engagement and nature and the economy.

Table 2.2 Overview of the provincial policy strategies selected for assessment during the evaluation.

Category	Strategies
Regular policy strategies	<ul style="list-style-type: none"> • Realising the Dutch Nature Network • Agricultural and private nature management • Improvement of water and soil conditions
Innovative policy strategies	<ul style="list-style-type: none"> • New Executional Arrangements • Invitational Nature • Facilitating Green Citizen Initiatives • System innovation Agricultural Nature Management • Nature Process Management • Nature-inclusive Agriculture

2.4.2 Phase 2 – Data collection and interpretation

The evaluation research began in the second phase. Data on provincial policy plans and the government frames was collected and jointly interpreted by researchers and the participants. For instance, interviews were conducted regarding experiences with implemented policies, collecting data on formal policy plans, assessing the current status of biodiversity, as well as case studies on innovative policy strategies and best practices.

'Bilateral consultations'

Noteworthy events during this phase were the 'bilateral consultations'. For the ecological analysis and the cost-effectiveness analysis, the evaluation researchers intended to use data on policy plans provided by BIJ12 and through document analysis, but this provided insufficient detail for a thorough and comparable analysis. In response, the researchers decided to visit each province to conduct in-depth interviews in order to gain more detailed information on the policy plans (including financial data and actual data on the current status of nature) and how to interpret these in relation to the province's aims as a means to further the assessment. The bilateral consultations were thus a type of interview between the researchers and provincial policymakers – for the sake of consistency we use the term 'bilateral consultation'.

Cost-effectiveness analysis

The limited availability of data also had repercussions on the cost-effectiveness analysis. The provinces were especially reluctant to supply such sensitive information since they were unclear about the exact purpose of the cost-effectiveness analysis. The researchers made several attempts to put cost-effectiveness on the agenda for discussion of the Workgroup, but the provinces held off.

A workshop with the Workgroup Nature Policy was organised to better explain the entire purpose of the reflexive evaluation, including the cost-effectiveness analysis and the matter of data collection. It was eventually decided to abandon the cost-effectiveness analysis – national and provincial governments decided it was too soon after decentralisation to reach any sound conclusions on the matter.

2.4.3 Phase 3 – Data analysis and shared interpretation

During the third phase, the evaluation researchers analysed the data. The analyses were subsequently validated and enriched during interactive group sessions with the relevant participants.

Group meetings on case studies

For instance, group meetings regarding the innovative policy strategies were organised in which the relevant social parties in each province reflected on the evaluation findings and their implications for the roles of the respective parties in implementing the strategy. Central focal points of the meetings were validating and enriching the conducted analyses and draw out lessons to inform the participants' learning processes in order to improve implementation of the innovative strategy.

Workshop Innovative Policy Strategies

During the collective workshop Innovative Policy Strategies all participants involved in the case-study research regarding the innovative strategies discussed the findings. The goal of the workshop was to identify the primary challenges and opportunities and their implications for the roles of the provinces

and national government, to draw lessons for policy practice. The workshop discussions also contributed to the design of the workshop Action Perspectives, which took place during phase 4.

Table 2.3 Information on the workshop Innovative Policy Strategies.

Workshop	Objective	Parallel sessions	Attendees
Innovative Policy Strategies (3/11/15)	Jointly identifying primary challenges and opportunities for broad implementation of the innovative strategies by discussing several cases for each strategy, with specific attention to the role of the province in this regard	Green citizens' Initiatives; Nature & Economy; Inviting Nature; Nature including agriculture; New executive arrangements; Process management	Commissioners EZ and IPO Provincial policymakers Societal partners (involved with studied cases)

Group-review sessions (only policymakers)

During the group-review sessions, the preliminary results of the model calculations on the effect of provincial policy strategies on the VHR-goals and the analysis of experiences with implementing these policies were presented to national and provincial policymakers. There were four sessions in total, each visited by representatives of three different provinces (and a representative of national government in two sessions). The sessions aimed to validate and enrich the analyses, and to list follow-up research questions. In addition, holding joint sessions (rather than individual, as was initially the plan) promoted the exchange of experiences to inspire policymakers' learning.

Sharing preliminary results with Deputies (BACVP)

Also during this phase, the preliminary findings were presented to the Deputies of the nature sectors in all provinces in a BACVP meeting in which participants jointly gave meaning to the findings and contributed to improving their framing.

2.4.4 Phase 4 – Formulating action perspectives and dissemination

Workshop Action perspectives

Finally, during the fourth phase, based on the evaluation conclusions, action perspectives and policy options were collectively formulated with national government, the provinces and societal partners. This took place in the collective workshop Action Perspectives. Based on the previous workshops and group-review sessions, and in conjunction with the Workgroup, specific policy themes were selected. All of these themes were perceived as posing challenges in attaining the three nature policy ambitions. In the workshop, preliminary analyses relating to these themes were shared, and respective challenges and requirements to tackle these were discussed in depth. Collectively, action perspectives for the various relevant actors to overcome the challenges were formulated and shared. This way, the workshop contributed to the concept version of the evaluation report, while also stimulating reflexivity on personal practice and learning.

Informal administrative meeting

During an informal meeting the State Secretary of EZ and the provincial Deputies invited the director of PBL to discuss the final evaluation conclusions. Action perspectives were formulated at this level of government, and experiences with nature policy since the decentralisation were shared.

Table 2.4 Information on the workshop Action Perspectives.

Workshop	Objective	Parallel sessions	Attendees
Action perspectives (6/9/16)	Jointly formulating action perspectives and policy options based on the findings of the Natuurpact evaluation, to produce input for the recommendations in the final evaluation report + stimulate ownership by the participants	Implementing and renewing policy instruments/self-realisation of nature; New executive arrangements; Justice equality and State aid; Water safety frameworks and floodplains; European Water Framework Directive; Improved environmental conditions; Societal initiative; Nature-inclusive agriculture	The provinces Representatives from EZ, IPO, BIJ12 and I&M Societal partners

Expert session

Based on feedback from the members of the Workgroup, a smaller session was organised to further sharpen the action perspectives and policy options. Some provincial policymakers felt the action perspectives proposed in the concept conclusions left too much to interpretation, and therefore provided several more sharply formulated suggestions.

Requested feedback

The evaluation researchers send the concept-versions of the final and background reports to the provinces, national government and involved societal partners to allow them opportunity to give feedback and check the reports on factual inaccuracies.

Formal presentation of final evaluation report (and background reports)

The final evaluation report was formally presented at a final symposium to which all involved participants and researchers were invited. The PBL director formally handed over the report to the State Secretary for EZ (Martijn van Dam) and the Deputy of Gelderland (Jan Jacob van Dijk). Table 2.2 reflects the reports that have been published on Natuurpact and its evaluation.

Table 2.5 Overview of the published reports on Natuurpact and its evaluation.

Reports	Dutch:	Authors:
Reflexive evaluation of the Natuurpact	Lerende evaluatie van het Natuurpact	PBL & WUR (2017)
Innovations of provincial nature policy in practice	De praktijk van vernieuwingen in het provinciaal natuurbeleid	Kuindersma et al. (2017)
Framing provincial nature policy	Het provinciaal natuurbeleid ingekaderd	Fontein et al. (2017)
Provincial nature policy potential in light of the European biodiversity goals	Potentiele bijdrage provinciaal natuurbeleid aan Europese biodiversiteitsdoelen	Van der Hoek et al. (2017)

Follow-up presentations

In response to the final publication, the researchers have received invitations from several provinces, EZ and a societal partner (and more are expected to follow) to present the evaluation findings. At these presentations, the findings relevant to the specific audience are highlighted and discussed with

policymakers involved with nature in the respective provinces. At the time of writing, these follow-up presentations were still taking place.

Finally, Table 2.3 presents an overview of the main evaluation events that stakeholders attended.

Table 2.6 Three key evaluation stakeholder groups and their participation in different evaluation phases, with the major evaluation-related activities set out by event.

Evaluation phases	1. Determining evaluation objectives and approach				2. Data collection			3. Data analysis & interpretation			4. Formulating Action perspectives & dissemination of final report		
	Preliminary interviews	LS I	LS II	LS III	Interviews	Bilateral consultation	Case studies	WS cost-effectiveness	WS New policy strategies	Group-review sessions	WS Action perspectives	Feedback on report	Expert session
Evaluation-related activities													
Provincial policy professionals (multiple administrative levels)	x	x	x	x	x	x	x	x	x	x	x	x	x
Commissioners (EZ and IPO)	x	x	x	x					x	x	x	x	x
Societal partners*	x		x	x	x		x		x		x		

* e.g. terrain management organisations, water boards, NGOs, agricultural representatives, private parties

3. Theoretical background

Our report ‘Characteristics of reflexive evaluation’ (Van Veen et al., 2016) provides a theoretical background on reflexive evaluation. The framework (shown in Figure 2.1) includes the expected outcomes and respective conditional factors used in this review. In this chapter we discuss these concepts and their operationalisation as they relate to this study.

3.1 The expected outcomes of reflexive evaluation

3.1.1 Knowledge-enriched policy practice

In our framework, we present increased policy impact as the ultimate outcome of reflexive evaluation. The introduction discussed the difficulties with assessing the policy impact of evaluation: this study occurs too soon after the evaluation to be conclusive, even if it were possible to establish a sound method for measuring impact. To establish increased policy impact, however, it may be possible to discern ‘intermediate’ outcomes. In light of this, the evaluation should lead to stakeholders’ knowledge-enriched practices in the area under investigation. By aligning the evaluation research to the policy practice being examined, the knowledge generated is expected to be optimally relevant to the policy process, i.e. optimally *usable*, facilitating timely adjustments and, ultimately, increased effectiveness at different levels in the policy arena.

The value of reflexive evaluation to its users

What does it mean for an evaluation to be ‘utilisation-focused’? In what sense can the process and outcomes of an evaluation be of value to policy practice? Various scholars have reflected upon these questions and have warned of an over-emphasis on the direct instrumental value of evaluation results – i.e. the assumption that the most important outcome of the evaluation is that it should lead to concrete actions taken on the basis of the information it has produced (see e.g. Kirkhart, 2000). Authors have emphasised the importance of unintended effects of evaluation, as well as the influence of conducting it, in the so-called *process use*, introduced by Patton in 1998. Here we will build on the framework constructed by Kirkhart (2000) on the different, what she calls, *influences* an evaluation might have⁶. The term influence may refer to any effect, impact or value the evaluation may achieve in a number of areas. We use the term ‘value’ to capture this.

We distinguish between five different values of an evaluation that, in our view, may be brought about by both its results (i.e. the findings presented in the final report and underlying background reports) and the process of the evaluation (e.g. interaction with and between stakeholders throughout the different phases of the evaluation)⁷. The five values of evaluation are conceptual, instrumental, strategic, affective and network, and will be briefly introduced below. Specifically, we hypothesised

⁶ While Kirkhart (2000) integrates three dimensions – sources of influence, intention, and time – we focus primarily on the different uses mentioned under ‘sources of influence’. In terms of the time dimension, this study took place during and immediately after the formal ending of the Natuurpact evaluation, positioning our findings in Kirkhart’s ‘immediate influence’ phase. In terms of the intention dimension, we do not make an explicit distinction between intentional and unintentional influences, taking both into account.

⁷ Hence, in terms of sources of influence, we do not follow Kirkhart’s distinction between results-based and process-based influence: in our view, both sources are more or less relevant to all uses and explicitly separating them leaves out this synergy.

conceptual and instrumental value to be of importance for contributing to a knowledge-enriched policy practice.

The *conceptual* value of evaluation occurs when an evaluation influences the way people think about, or understand, a given area, or a policy programme, *'without any immediate new decisions being made about the program'* (Bayley 2008:2). Conceptual use is sometimes also referred to as *'enlightenment'* or *'demystification'* (Kirkhart, 2000:9), or as the *'cognitive dimension of process use'* if the changed understanding is a result of the discussions and reflections that are part of the evaluation process (Kirkhart, 2000:10). This conceptual use of evaluation may also be understood as learning as it entails an increased – or changed – understanding of the issue at hand (Verwoerd, 2016). This conceptual value may transform into *instrumental* value of evaluation: *'direct, visible action taken based on evaluation findings'* (Kirkhart, 2000:9, referring to Rich, 1977), such as a policy change or the implementation of recommendations. The *strategic* value of evaluation refers to the role an evaluation may play in advocacy, argument and political debate (Kirkhart, 2000). Sometimes strategic value is meant in the sense of *'symbolic'*, e.g. if the evaluation is used to justify decisions already made, or to postpone decisions, or in the sense of *'persuasive'*, e.g. if the evaluation is used to convince others to support a decision. Strategic value can also refer to using the evaluation to draw attention to certain problems or shifting the discourse. The *affective* dimension of evaluation value refers to *'the individual and collective feelings of worth and value about themselves, the evaluation or the program, that results from the evaluation process'* (Kirkhart, 2000:10 referring to Greene, 1988b). Finally, we introduce the concept of *network* value to refer to the value of the evaluation in terms of building networks and strengthening relationships. As the evaluation took place in a multi-stakeholder context, with evaluation activities that brought together diverse stakeholders in learning sessions, we anticipate that this value will emerge from our study. The results on the manifestation of these values are discussed in chapter 5.



Bird-eye view of the Netherlands

3.1.2 Multi-directional accountability

Horizontal accountability

In our framework, we discuss how reflexive evaluation is capable of bridging the supposed dichotomy between evaluation for accountability purposes and evaluation to support learning. Scholars have explored how these two concepts relate to each other and found that they are not as different from each other as generally assumed. First, accountability comprises more than just vertical accountability (towards a funder or commissioner; Richmond, Mook, & Quarter, 2003). Any policymaker is also accountable to stakeholders who are either affected by or involved with developing and implementing the policy. The provinces are thus accountable to societal partners, businesses and citizens in taking into account and safeguarding their interests in policy practice to the best of their abilities. Naturally, vertical accountability is not neglected but is subsumed in horizontal accountability, where the commissioner holds a more equal position in relation to stakeholders affected by the policy. In other words, rather than one formal commissioner, *all* relevant stakeholders are recognised as equally important commissioners who all have specific perspectives and values that should be taken into account when developing and implementing the policy. Reflexive evaluation is proposed to support horizontal accountability by facilitating familiarisation with the perspectives and interests of relevant stakeholders, as a first step in developing policy that better aligns to them.

Learning (internal accountability)

Furthermore, we explain how policy professionals are accountable for attaining their own mission, which is termed internal accountability (Ebrahim, 2005). Internal accountability conceptually aligns to learning. Especially in complex contexts that are subject to unpredictable change, reflexive evaluation may serve to optimise learning processes to better respond to changes and increase policy impact and goal attainment. With a better understanding of policy theory and processes that influence policy execution there is a greater likelihood of goals being attained and, consequently, internal accountability. In this conceptualisation, we relate internal accountability to the expected conceptual and instrumental use of the evaluation, discussed in the previous paragraph.

Following this line of reasoning, accountability and learning are reconciled through the concept of multi-directional accountability.

3.1.3 The evaluation leads to societally robust knowledge

Another outcome concept that was introduced in Van Veen et al. (2016) is that of societally robust knowledge. We expect (policy) practices to become more informed and enriched by the knowledge generated through the evaluation process. Likewise, we expect the research process to become more practice-informed, resulting in knowledge that is enriched and societally robust while remaining scientifically sound. It means that the knowledge generated is highly contextualised, as the process of its production occurs in intense interaction between the stakeholders, in the context of its application. Societally robust knowledge (Nowotny, 1999; 2000) is context-appropriate, broadly supported and sustainable. This implies that if participants consider that the outcomes of the evaluation are appropriate to their context, the knowledge generated can be considered societally robust. Which brings us back to knowledge enriched policy practice, and the different ways the Natuurpact evaluation is of value to its participants.

3.2 Conditional factors to realise the expected outcomes

Furthermore, in our framework we discuss conditional factors required to establish the expected outcomes of reflexive evaluation. Assessing the quality of the evaluation is exceedingly difficult, as it is nearly impossible to predetermine a rule for the quality of reflexive evaluation (Belcher, Rasmussen, Kemshaw, & Zornes, 2016; Shulha, Whitmore, Cousins, Gilbert, & al Hudib, 2015). To this end, authors focused on formulating conditional factors to guide the research and establish the expected outcomes.

3.2.1 Stakeholder selection

We proposed that at the start of the evaluation it is necessary to consider which stakeholders to involve. Including a variety of stakeholders during the evaluation contributes to increased understanding, legitimacy and commitment to conduct an evaluation that is actually used to inform practice. As the group of relevant stakeholders is potentially large for nature policy, it is legitimate to be pragmatic in selecting stakeholders: scholars agree that at least the primary intended users of the evaluation findings should be involved, while other parties should remain informed on the evaluation and its progress (Patton, 2008).

3.2.2 Process of reflexive evaluation

We also discuss several process requirements of reflexive evaluation: stakeholder engagement and alignment between policy practice and the evaluation research.

Stakeholder engagement

Reflexive evaluation is a participative approach in which stakeholders have an active role in shaping and conducting the evaluation research. It is therefore important the participants are willing to participate. This may be established by encouraging their ownership and commitment by aligning the evaluation research to the participants' practices (discussed below) and by a sense of urgency and a need for change. Moreover, for the participants to be engaged it is vital that there are levels of mutual trust between them and the researchers in order to ensure they share relevant information required for the evaluation. A transparent research process supports mutual trust.

Table 3.1 The level of involvement of participants during traditional evaluation approach and reflexive evaluation

	Level of involvement of participants →	Low (traditional evaluation approaches)	High (reflexive evaluation)
Research phase ↓			
Phase 1: Determine evaluation objectives and methods		Determined by researchers (methods)	Participants are consulted (evaluation objectives) Jointly determined
Phase 2: Data collection		Participants are not involved, passively provide data	Participants are consulted, actively provide data Data is jointly collected
Phase 3: Data analysis and interpretation of the findings		Receive results	Give feedback on results Give feedback, share interpretation and draw shared conclusions
Phase 4: Dissemination of the evaluation findings		Receive report (researchers are sender)	Vision of participants is visible (e.g. case stories), researchers are sender Jointly written report (both are senders)

Alignment between policy practice and evaluation research

Active involvement during each evaluation phase

During the evaluation, the participants and researchers collaborate. The participants are actively involved with each evaluation phase and own the process. The right column in Table 3.1 shows the level of involvement for reflexive evaluation as proposed in the literature, while the left column shows the involvement of participants in more traditional approaches. In reflexive evaluation, participants jointly determine the evaluation research questions and the methods used to answer these. They jointly collect the relevant data and give feedback on, and participate in, a process of shared interpretation and drawing conclusions based on the findings. Finally, the final evaluation report is written collaboratively and is a joint product of evaluators and evaluated, displaying the shared ownership of both parties.

Continuous cycle of reflection and adaptation

To support sustained participation of the participants and alignment of the evaluation research to their practice, there need to be frequent moments of interaction to reflect on different aspects of the evaluation (e.g. its process, progress, findings) (Edelenbos & van Buuren, 2005; Preskill & Torres, 2000). During such moments, researchers monitor the alignment to the informational needs of the participants and adapt the evaluation design as appropriate.

Interactions designed to aid mutual understanding

The purpose of these frequent interactions is also to encourage learning processes based on the (preliminary) findings of the evaluation (Van der Meer & Edelenbos, 2006), but also based on interactive frame reflection with other stakeholders. Recognition of other underlying values regarding the policy under scrutiny is said to allow for more effective collaboration through better understanding of each other's frames (Schön & Rein, 1994).

3.2.3 Conceptual difference between outcomes and factors

Before we continue with methodology, we briefly draw attention to the difference between outcomes and factors that contribute to them. Though these seem clearly conceptually distinct, in analysing the data we found concepts that could simultaneously be considered an outcome as well as a factor for a different outcome. During process evaluation it is inevitable that what may be perceived as an intermediate outcome also serves to manifest outcomes that play out on a longer timescale. For instance, increased mutual trust may be an outcome of interactive processes, but may also be perceived as a factor for improved collaboration. We acknowledge this conceptual ambivalence and have structured our results section according to the anticipated outcomes and factors as proposed in this chapter.

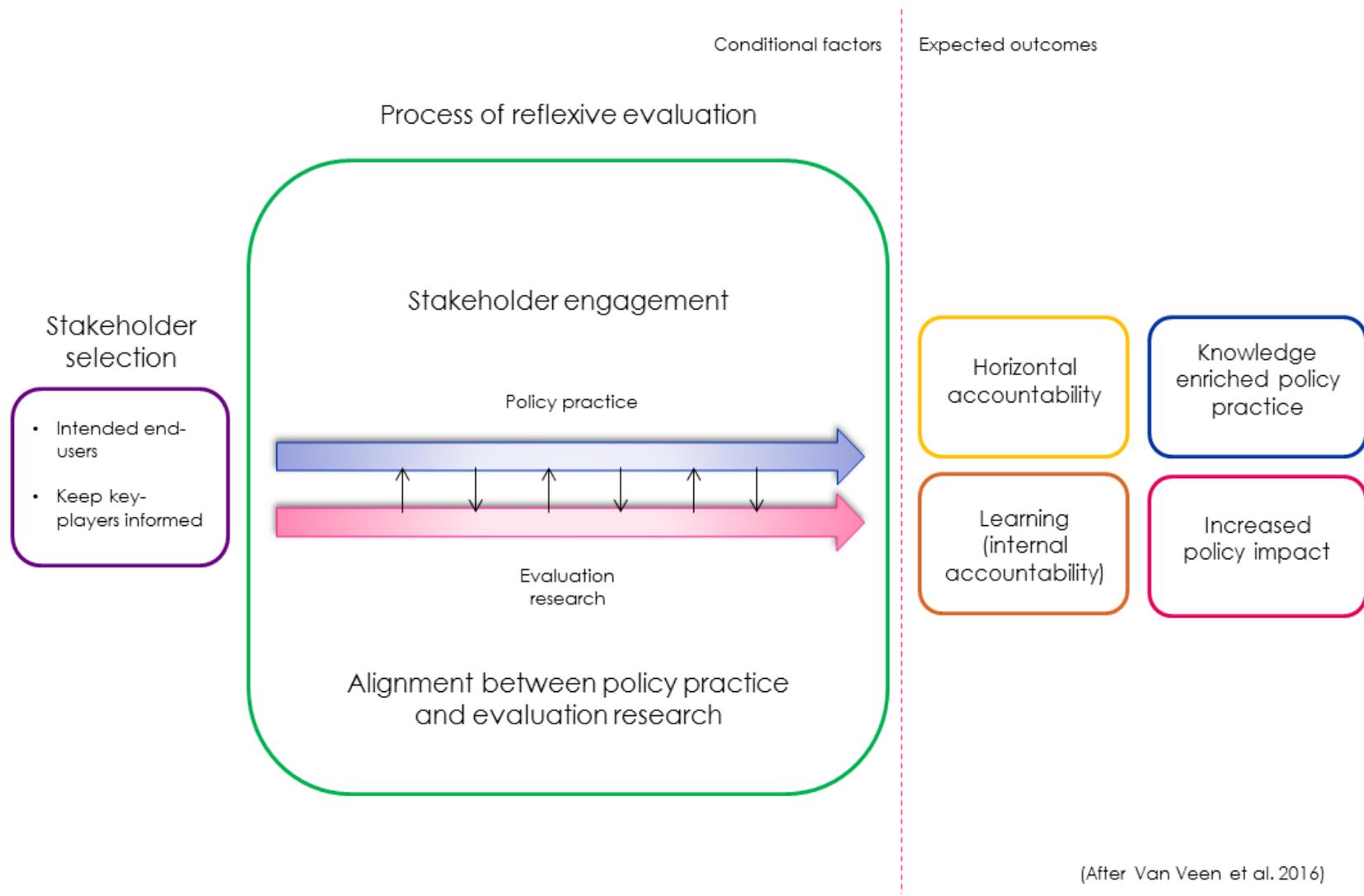


Figure 3.1 Theoretical framework after Van Veen et al. (2016) illustrating the conditional factors and the expected outcomes characteristic of reflexive evaluation.

4. Methodology

In this chapter, we discuss the methods used to answer the two main research questions on the value of reflexive evaluation for policy practice, and factors that contribute to this. We explain the different data sources that were used (transcripts, observations and (group)interviews), how respondents were selected and data were obtained, and how the analyses were subsequently conducted. We also highlight the analysis of so-called ‘learning questions’ and the ‘tough issues’ we derived from these, in relation to the ability of the reflexive evaluation to evoke reflexivity as part of research question one.

4.1 Data collection and analysis

4.1.1 Non-participant observation

VU researchers attended a range of collective workshops and group sessions (see Table 4.1). Frequently at least one of the VU researchers would facilitate the joint discussion. Another would make non-participant observations, recording actions and reactions of the participants as they occurred, and, if relevant, the response of evaluation researchers to these. Observation was systematic, focusing on questions that were asked, discussions that arose and how these were dealt with by participants and evaluation researchers (e.g. ignored, satisfactory concluded, put on (evaluation) agenda). There was also a focus on the demeanour of participants: were they sharing openly, or did they make defensive remarks, such as comments that suggested distrust or reluctance to participate, etc.

Table 4.1 Overview of attended evaluation events during which VU researchers observed (inter)actions of intended end-users and evaluation researchers.

Evaluation events attended by VU researcher(s)

Collective learning sessions, e.g. workshops	All three learning sessions prior to development of evaluation plan
	Workshop CEA
	Group meetings in light of innovative policy strategies study
	Group-review sessions in light of ecological analysis (including citizen initiatives)
	Workshop Innovative Provincial Policy Strategies
	Workshop Action Perspectives
	Follow-up expert session on action perspectives
Evaluation research related interactions	Interviews in light of innovative policy strategies study (including citizen initiatives)
	10 personal interviews with provincial policymakers and societal partners on fail and success factors with implementing policy strategies
	1 ‘Bilateral consultation’ in light of the CEA and ecological analysis

4.1.2 Secondary data

Not all research-related interactions could be attended. For instance, the collective workshops often comprised too many parallel sessions for the capacity to attend them all. Nor was it possible to be present at all evaluation research-related interactions such as interviews with policymakers and societal partners and the 12 ‘bilateral consultations’. Sometimes there was enough time, at others it was thought inappropriate for the VU to attend due to the sensitive information under discussion. Therefore, when available, the audio-recordings, transcripts, and summaries of the interactions were used for analysis. These were analysed as the systemic observations, with a focus on (re)actions and demeanour of the intended end-users, types of questions and comments, which discussions arose, etc.

4.1.3 Interviews

Interviews with participants and evaluation researchers

To gain insight into how the evaluation has been of value to its participants’ practice and which factors were important in achieving this, we conducted several consecutive rounds of interviews, including a focus group discussion with the Workgroup Nature Policy. The interviews sought to discuss the different ways the evaluation is considered valuable, as well as to test assumptions (both on types of value and on factors) emerging from the theoretical framework on reflexive evaluation (Van Veen et al., 2016). The interviews were therefore of a semi-structured character, with a specific focus but also allowing sufficient room and space for emerging issues the respondents thought of as important.

Table 4.2 Three rounds of interviews and respondents.

Interview rounds	Respondents
Round 1	<ul style="list-style-type: none">- 1 provincial policymaker (part of Workgroup, and representative of Workgroup in closer interaction with evaluation researchers)- 3 provincial policymakers (not in Workgroup, but attended at least 2 evaluation events)- 1 societal partner- 1 representative of EZ- 1 representative of IPO
Round 2	<ul style="list-style-type: none">- Members of the Workgroup Nature Policy (focus group discussion)
Round 3	<ul style="list-style-type: none">- 2 representatives of EZ- 1 representative of IPO- 1 member of the Workgroup (who was not able to attend the focus group session)

A total of 11 interviews were held, and a focus group session with the Workgroup Nature Policy (11 members and chairperson) in three consecutive rounds. Table 4.2 shows these rounds and the respondents. For the first, more explorative, round, we selected provincial policymakers who had a moderately active role in the evaluation, i.e. they had attended at least two workshops, and had

shown either a strong positive, neutral or more negative attitude towards the evaluation and PBL (with the exception of one provincial policymaker who has a highly active role as representative of the Workgroup to test-run our first interview). We also interviewed a representative of a societal partner (selection criteria: > 2 evaluation events attended), and representatives of EZ (policymaker) and IPO (advisor) (selection criteria: > 3 evaluation events attended). For the second round we conducted a focus group session with the members of the Workgroup Nature Policy. The third comprised a final validation round with two representatives of EZ (a manager and a policymaker) and one from IPO (manager).

Also, eight interviews were conducted with the evaluation researchers (both PBL and WUR) throughout the course of the evaluation. Of these, three were held with the project leader (PBL) and the *ad interim* project leader (WUR) at the start of the evaluation, halfway through and after publication of the final report. During the interviews with the evaluation researchers their view on the participation of the intended end-users was discussed, the progress of the evaluation and the dilemmas they faced in implementing process requirements and dealing with contextual factors. These interviews were also semi-structured.

Interview design

The design of the interviews with the intended end-users from the first round was informed by the findings from the secondary data analysis and the observations, and the sensitising concepts from the theoretical framework. During these interviews the various ways the evaluation was found valuable were discussed, as well as specific factors contributed to this view. The role of the evaluation researchers was also discussed, as contextual factors that influenced how the evaluation was conducted.

The findings from the first round were used to inform the design of the focus group in the second round. In the focus group, openly ways the evaluation was experienced to be of value were first listed and discussed. In the second part of the focus group, we asked the respondents to prioritise 15 different factors that they felt had played a significant role in realising the evaluation's value for their practice. These factors and the role of the researchers on manifesting these were collectively discussed.

Finally, the interviews in round three served as final round of enriching and validating. In addition, by allowing the different rounds to inform each other, we also confronted the perspectives of the provinces and of the commissioners (EZ and IPO) with each other.

A further important source of data were the bi-weekly meetings of the evaluation researchers. At these meetings, the process and progress of the evaluation were discussed, and reflections on the participation of the intended end-users, concerns of the researchers, dilemmas they faced, etc., were shared. These meetings, as well as informal discussions, proved a valuable source of information to better interpret our findings in relation to their context.

All interviews were, with permission of the respondents, audio-recorded and transcribed. Analyses were made by coding, using MAXQDA 11 (11.0.7). Coding was open, but we were sensitive to concepts derived from the theoretical framework discussed in Van Veen et al. (2016).

4.2 Learning questions analysis

We hypothesised the evaluation would result in conceptual and instrumental value through alignment of the content of the evaluation to the informational needs of its participants. To critically review not only the success of the Natuurpact evaluation in realising this alignment, but also potential shortcomings in this regard, we have retrospectively analysed so-called learning questions of the provinces, to allow reflection on the alignment between the evaluation scope and the informational needs of its participants. As it comprised a rather different approach from the analysis discussed in the previous paragraph, we here highlight how we conducted this analysis, also providing a brief description of the underlying theory.

4.2.1 Dynamic Learning Agenda theory

The Dynamic Learning Agenda (DLA) is a tool to capture and inform emergent learning processes. In this study, we adapted this tool and applied it retrospectively, to construct narratives of learning processes. According to DLA theory, questions raised (we term them learning ‘questions’, but in fact *all* comments that express a concern or struggle relating to the topic under scrutiny are relevant) are understood as a reflection of challenges or issues an actor perceives in taking action towards a more desired state (e.g. realising certain policy goals) at a given moment (Regeer, Hoes, van Amstel-van Saane, Caron-Flinterman, & Bunders, 2009). Such questions may be monitored on a Learning Agenda; are they solved, or do they remain on the agenda for a longer period of time? In the latter case, we speak of persistent learning questions or ‘tough issues’ (Van Veen, de Wildt-liesveld, Bunders, & Regeer, 2014).

4.2.2 Inventory of learning questions

All available transcripts – for example, from collective sessions and interviews for evaluation data collection – were analysed for learning questions. Our inventory comprised over 450 learning questions. Learning questions were recognised if they related to difficulties the participants perceive in developing and implementing nature policy (not necessarily phrased as questions), e.g. in relation to government legislative frameworks, collaborations with societal partners, conflicts with farmers, etc. Questions were excluded when they concerned general informative questions (e.g. *‘Until what time will this session last?’* However, we did not develop elaborate exclusion criteria, as we believe that even an ostensibly general question may concern a tough issue. For instance, the question *‘What is the national government legislation with regard to the national waters (Dutch: Rijkswateren)?’* may seem to be of a general informative nature, but in our analysis it kept returning in different forms and appeared to signal a collective knowledge gap on the new roles and tasks of national government following the decentralisation.

4.2.3 Analysis of learning questions

All learning questions were first clustered thematically through open coding. Themes included, for instance, ‘Policy strategies’, ‘Working with societal partners’ and ‘Role of provincial government’, to name a few. Applicable sub-codes were added. For instance, ‘Policy strategies’ gained two sub-codes: ‘The effectiveness of policy strategies’ and ‘Implementing policy strategies’. Next, the learning questions were coded by theme according to the evaluation event in which they were posed. Finally, we analysed the type of knowledge the learning questions called for, for which we followed Pohl and Hirsch Hadorn (2008): system knowledge, target knowledge and transformation knowledge.

The analysis provided an overview of different thematic clusters containing learning questions on different knowledge types, posed over a timeframe of three years. The analysis was validated during the three rounds of interviews (see 4.1.3).

4.3 Research reliability and validity

As already touched upon several times, the VU researchers who conducted this review also played an important role in shaping the reflexive evaluation. We anticipated that researcher bias might (subconsciously) affect our interpretation of the findings. To guarantee the scientific rigour of this study, we took several steps. First, we expanded our research team with researchers who functioned at a greater distance from the reflexive evaluation, and with whom we interactively reflected on interpretation to reduce inadvertent bias. Furthermore, we obtained data through different methods, which were confronted with one another (triangulation). The respondents validated our findings by discussing them during the three rounds of interviews. Validation also occurred by evaluation researchers, with whom we frequently discussed our progress and their feedback. Through validation we guarded against blind spots and misinterpretation of meaning. We also ensured researcher triangulation by conducting the analyses individually and discussing the differences in data interpretations before integrating the analyses. Finally, scholars in the fields of policy science, public administration and environmental governance (with experience in multi-actor multi-level governance processes and policy evaluation) conducted an external review of our work.

5. The value of the Natuurpact evaluation

In this chapter we present the different ways the participants perceive the evaluation (both its process and its results) has been of value to their practice, based on the interviews and focus group we held (5.1). In paragraph 5.2 we discuss the results from our analysis of the learning questions and present three categories of informational needs that are currently unaddressed in the context of the evaluation. Furthermore, we reflect on how the manifested values relate to the expected outcomes based on our theoretical framework discussed in chapter 3.

5.1 Five different value-types

In our analysis we observed five different types of value: network, affective, conceptual, instrumental and strategic value. The participants mostly spoke first about the network value of the evaluation, in appreciation of all interactive sessions with a small number of other provinces or the entire network around the Natuurpact. As part of this, but analytically distinct, was the affective value of the evaluation – building trust and also a sense of togetherness throughout the process. Second, the conceptual value, in terms of the knowledge generated and understandings gained, was important to all participants. Furthermore, the strategic value of the evaluation is perceived as highly important for participants from all levels of government involved. And finally, we identified examples of instrumental value. All values are explored in greater depth, and illustrated with examples, below.

5.1.1 Network value of the evaluation

The value of the evaluation for their network was recognised by the majority of the participants. The process of evaluation contributed importantly to fostering contacts and interactions between different stakeholder groups. They speak about expanding their network and strengthening existing relationships.



Provincial policymakers exchanging experiences during a collective workshop

Network expansion

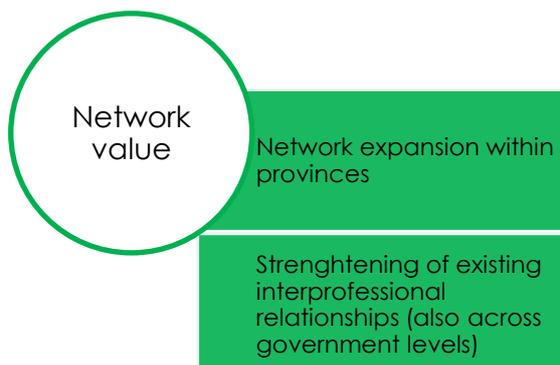
The participants indicate that the evaluation workshops allowed them to become acquainted with parties they would not otherwise have encountered, such as knowledge institutes or societal partners; it led to network expansion to benefit policy practice in their province. A provincial policymaker remembers: *'It was during this workshop on agriculture when I learned Campina (Dutch dairy cooperation) was also employing nature-inclusive agricultural methods. Then I thought, well, perhaps we should also get in touch with them...'* (P1)

Strengthening existing inter-professional relations

Participants further agree that the evaluation has had a positive effect on strengthening existing inter-professional relations. From the provinces, a participant states: *'It's about getting to know each other. That also counts for our colleagues. And indeed also for environmental organisations, colleagues from the Ministry of Economic Affairs... It's about getting to know those parties you have to deal with in the process [of nature policy].'* (P3) Also, colleagues between the provinces, especially within the workgroup Nature Policy, say they have become better acquainted with one another and so have a better idea of each other's value, which has a positive effect on their collaborative capacity. This can, however, not be fully ascribed to the evaluation: the Workgroup meets monthly for other tasks concerning nature policy.

One specific relationship that was frequently discussed was that between provinces and the Ministry of Economic Affairs. Participants suggested that the two parties have a history of distrust, which was further complicated by the decentralisation and the consequent redistribution of tasks and responsibilities. When asked how the evaluation has contributed to this relationship, provincial policymakers responded that the dialogue has progressed significantly over the last few years, as the following quotes illustrate: *'Last year I also spoke with EZ. And you notice that – on the base of increased trust – there's more opening for discussion compared to five years ago. And that's great.'* (P3) A representative of the Ministry of Economic Affairs reflects: *'It has transformed from a state of distrust towards EZ, towards an interest in getting together.'* (C1) A policymaker comments: *'At some point, EZ has to deal with Europe. The findings in the report, in case a province doesn't achieve the results on biodiversity, can have consequences. But to be afraid of these consequences, that would be a sign of mistrust.'* (P4)

Some representatives of the provinces also indicated that the Ministry's participation in the evaluation was too limited, which might have impeded building trust. Nevertheless, participants concurred that there have been small steps towards each other, illustrating how the evaluation has had effect on relations across different levels of government.



5.1.2 Affective value of the evaluation

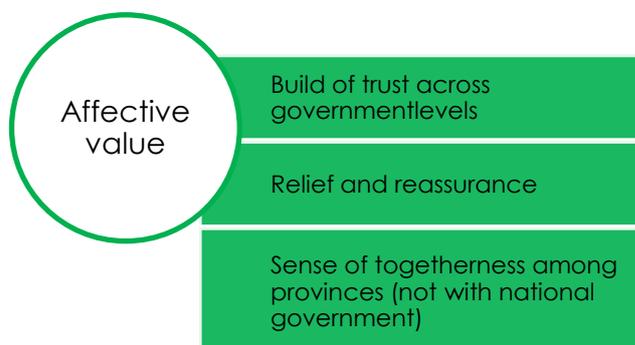
When further discussing the value of an expanded and strengthened network, the participants commented on the affective value of evaluation. For instance, the strengthening of the relationship with EZ originates from increased trust resulting from more frequent interaction in the context of the evaluation.

Relief and reassurance

The decentralisation of nature policy implied a major transition for the provinces. Knowing how other provinces are doing created a welcome sense of relief: *‘As a result of all the sessions and also the final report I think “well, we are not doing too bad as a province”. Earlier I was afraid we had missed opportunities, but now I feel quite relieved. Not so much regarding biodiversity, but also the new policy strategies.’* (P4) Recognising that their colleagues are dealing with the same difficulties – primarily in taking on their new role, and shaping policy for societal engagement – was reassuring and seemed to legitimise their own struggles. As one participant illustrates: *‘Before the evaluation, there were two main themes I found challenging: nature and economy, and nature and society. [...] I’ve found that, luckily, these themes are still in their infancy in all provinces.’* (P15)

Sense of togetherness

For some, ascertaining that the provinces encounter the same difficulties and are all looking for methods to best deal with these, also inspired a sense of togetherness: *‘There is a lot to be done, but at least we’re not alone in this.’* (P15) This sense of togetherness is also described by other provincial policymakers: *‘It creates a sense of “we’re in this together”. It makes you aware that you share a responsibility.’* (P14) Not everyone shared this idea, however. Other provincial policymakers are more critical of the role of the evaluation regarding this feeling of togetherness. For instance, one provincial policymaker comments: *‘This sense of togetherness... It exists, and the evaluation contributes to it. But it’s not the sole source, I think that’s very important to underline.’* (P3) Togetherness does not seem to include EZ and IPO or other parties, with which a commissioner from EZ concurred: *‘The idea of evaluating collectively is appealing, because you do hope that after a few years a kind of community comes to life, you share a process and you see each other frequently... [...] But I don’t see that that has happened with EZ yet. Our role was too limited during this evaluation to inspire this sense of togetherness.’* (C1)



5.1.3 Conceptual value of the evaluation

Furthermore, the provinces discuss the knowledge gained and understanding generated as a result of the evaluation – its conceptual value. We distinguish three dimensions: the first relates to single-loop learning (Argyris & Schön, 1978) that one would expect of any programme or policy evaluation (e.g. knowledge on where we stand, on strategies and instruments to obtain goals and deeper

understanding of underlying issues that stand in the way). The second dimension is more characteristic of participative evaluation and pertains to double-loop learning: learning about personal and others' perspectives and values, and underlying assumptions that determine one's courses of action (ibid.). Finally, the third dimension is experienced at a more collective level and concerns increased conceptual coherence and the shared ambition of nature policy.



Policymakers during the workshop Action Perspectives

Single-loop learning

Current state

The evaluation has given insight into where the Netherlands stands regarding biodiversity targets. While some view this status positively, others are more critical: *'For me the most important conclusion is that, despite all that we are doing, it is not enough.'* (P14) Regardless of how the status is appreciated, the evaluation has provided the knowledge needed to assess current progress towards international goals. The same was true regarding the current state of provincial biodiversity; the evaluation made it possible to form a better image of a province's position in relation to the others. Though the provinces were explicit in wanting to steer clear of negative benchmarking, comparison between them is found valuable and informative. One policymaker stresses: *'It helps to get a better idea of your own position as a province within the total spectrum, whether you do a lot or a little.'* (P9)

Overall, there seemed to be broad satisfaction with the evaluation scope. The provinces are content with the knowledge on the current state of Dutch nature the evaluation provides. Some, however, did comment on the evaluation scope during the workshops. They, for instance, questioned the focus of the evaluation on the VHR and KRW and as a result the limited insights into nature outside the designated areas (Natura2000 and the Nature Network). As a result of such remarks, it is being contemplated to broaden the scope of the evaluation in the following period.

Increased knowledge about variety of strategies

Policymakers speak of feeling inspired by other provinces on how to approach nature policy. *'The inspiration and recognition lead to a number of eye-openers. That was really beneficial.'* (P9) A

concrete example of an eye-opener concerned the realisation of the Nature Network: *'I realised that in our province, we approach realising the Nature Network in a very conservative, old-fashioned way. I learned that there are many other way to realise these goals. That diversity within the Nature Network can be achieved through other means than buying land and turning it into nature. That realisation was very valuable to me.'* (P3) Another concrete example concerned inter-sectorial collaboration: *'In our province we approached the collaboration between nature and water in a sectorial way. Two sectorial administrative agreements were signed, whereas in another province they integrated this from the start. Which seemed much more logical to me. It was interesting and inspiring to see that other provinces indeed approached it differently.'* (P9) And another participant adds: *'It was very inspiring to hear how another province was approaching green citizen initiatives. They had been working on this for a while and could talk it about very excitingly. This is a topic that we have not yet engaged in, and through these inspirational presentations you immediately get ideas and think: "Hey, that works! We could also do it like that."'* (P13)

Listening to the experiences of other provinces also made clear the differences between them, which was perceived as useful. One participant says: *'Not only recognition, but also sheer astonishment about the variation between the provinces. That is also very informative, decentralising on the same theme and finding so many different approaches.'* (P12) The differences between the provinces – e.g. geographical but also organisational – as explanatory factors for the diversity in implemented strategies and progress was enlightening: *'It's also very inspiring to think about provinces who have less to spend and have more difficult circumstances, and how they accomplish what they do.'* (P9)

Double-loop learning

In-depth understanding

Members of the Workgroup indicated that the evaluation provided opportunities to have more profound discussions on nature policy which in the hectic day-to-day schedule does not easily take place. This led to deeper understanding of their own assumptions on policy theory and increased mutual understanding. *'Within the workgroup Nature policy we usually talk at a process level, quite technical. The evaluation gave space to delve more into the content, to reach more depth.'* (P11) Similarly, the group sessions added value: *'I did sometimes feel "yet another session I have to go to..." But then, when you get there, and you are all together again, there is room to take some distance and reflect.'* (P11) And another provincial policymaker adds: *'We have been talking about the added value of form [value of interactive sessions] but we should also mention the content. The evaluation contribution is also in the topics it addresses, which can form input for policy processes. It helps structure discussions.'* (P14) The evaluation thus allowed for provinces to take a step back and increase their understanding of the bigger picture of the system in which they operate.

Reflection on underlying assumptions

Understanding was not only gained of the current status, but also of the implementation of provincial strategies. Discussing implementation and the rationale behind certain decisions made some provinces more conscious of their assumed policy theories and possible different perspectives on this. *'We discussed things like "Why did we approach it like this?, did it work?" We really went in much deeper than we ourselves ever would. Of the entire evaluation I thought this was the most evaluative.'* as a policymaker(P2) reflects on the bilateral consultations.

Frame reflection

Gaining insight into the perspectives of different parties on provincial nature policy and how these policies affect their practice was perceived as a valuable attribute of the evaluation. Such parties include, for instance, private parties, as a provincial policymaker comments: *'Private parties commented on aspects they run into with nature policy. It made me think "Okay, I hadn't considered that." So it provides you with a different lens than just the administrative one.'* (P2), or, put differently, it led to frame reflection (Schön & Rein, 1994). This was viewed as a valuable quality of the evaluation, as on a day-to-day basis there is little space for exploring other actors' perspectives.

Conceptual coherence

We observed an increase in mutual understanding and a shared language among the participants: *'The benefit of getting to know each other is that you learn to speak the same language and can more easily connect. We understand what we talk about, this makes it easier.'* (P3) In their interactions, we observed an alignment of language use between national and provincial policymakers and evaluation researchers that helps communication. Shared terminology, for instance, includes terms such *policy innovations*, to discuss new strategies to further nature policy, and *policy plan potential* and *executorial potential of policy plans*, marking the difference between written plans, and the potential of those plans in terms of attaining biodiversity goals, taking into account contextual factors.

Shared ambition

Relating to the sense of togetherness we observed as part of affective value, the broadened ambition of nature policy seemed more shared among the participants, especially in the Workgroup. In the Natuurpact it is stated that the ambition of nature policy is to be broadened and to better embed nature in society at large. In determining the evaluation plan, this broadened ambition was formulated in terms of three equal nature policy goals: increasing biodiversity, strengthening the relation between the economy and nature, and increasing societal involvement with nature.

The members of the Workgroup commented that frequent discussion on the three ambitions contributed to a shared understanding of the importance of a broadened outlook on the ambitions. This notion is more accepted by the Workgroup and more internalised as a goal to strive to achieve. When reflecting on this change, a provincial policymaker explained that it took some time to accept the divergence from the prevailing unilateral focus on biodiversity: *'In the beginning I just expected these other goals would be positioned as subordinate and conditional for biodiversity. Of course, it's fine to make them into solitary goals, it's just rather different than how we've always done it and I didn't expect it.'* (P12)

We also see indications of appropriation of the broadened scope of the ambitions at higher administrative levels. For instance, a policymaker spoke of a Deputy who stressed the importance of societal engagement throughout his professional network. This was valuable for increasing the recognition of the importance of the broadened take on the ambitions by their administration: *'Our deputy has really taken on the recommendations [in the final report, concerning the two new ambitions], which is really good as we all have to deal with this together. He also communicates to his colleague-deputies: "What actions should we take?" This also legitimises the actions we [the Workgroup] take in response to the recommendations.'* (P5) This change in policy discourse was also apparent during an informal administrative meeting (which included representatives from societal

partners, provincial deputies, representatives from EZ and from PBL). Most of those present recognised societal engagement as an end in itself.

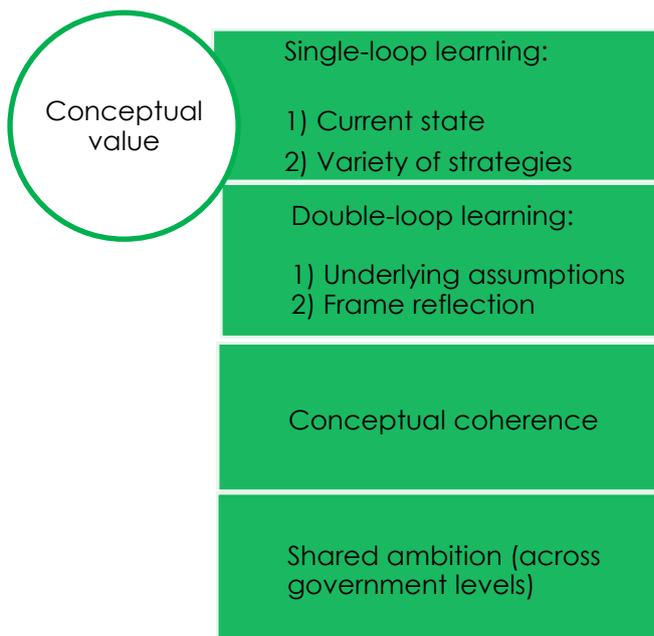
We thus see the broadened ambition become more shared among policy professionals (across government levels) during the course of the Natuurpact evaluation. The new ambitions have gained traction and are increasingly recognised as ends in themselves, rather than a means to a biodiversity end.

Limitations to conceptual value

Interestingly, most of the conceptual value seems to be limited to provincial policymakers who have been more active in the evaluation process: *'Our [members of the Workgroup] understandings have converged, we have gone to a process together during which we frequently had opportunity for discussion. I notice that my in-house colleagues have not always made that same step as they have been less part of it.'* (P2)

Also, commissioners from the Ministry indicate that much of the learning was focused on issues faced by the provinces. *'As a ministry, we also have learning questions regarding the decentralisation process. These were not addressed in the evaluation, as, of course, the emphasis of the decentralisation was with the provinces. Now, we are aware of this, we will bring our learning questions in in the next phase of the evaluation.'* (C3) At the same time the Ministry indicates that the evaluation has really aided the process of decentralisation because it brought the parties together in a learning process: *'The value of mutual learning has become clear to the provinces, because of the evaluation.'* (C3)

As touched upon in the introduction of this chapter, in paragraph 5.2 we return to the conceptual value of the evaluation in discussing informational needs articulated by the participants (in the form of learning questions) so far *not* addressed by the evaluation.



5.1.4 Instrumental value of the evaluation

If translated into concrete actions and decisions, the conceptual use of the evaluation acquires instrumental use. As we discuss in our introduction, this review took place soon after the publication of the final report. As little time has passed since then, it is unlikely that action taken on the basis of the evaluation is yet visible – rendering fully determining instrumental value impossible. Nevertheless, despite the short time span, we do see some clear examples of how the evaluation is used and implemented in policy practice.

Supplementation and informing Nature Visions

A prominent example is the use of the evaluation outcomes in the Nature Visions being formulated by the provinces. These visions describe the provincial plans for nature policy for the coming years. The analysis shows that some provinces used the evaluation findings to inform and supplement the development of their Nature Visions. *'In our Nature Vision we have discerned four parts. First, where are we now and where do want to go. Second, what are the citizens' desires regarding nature. Third, what value can we add to nature. And fourth, what do we do outside of the nature network. Yesterday, we had a meeting with partners about the first part and we have said, let us first consider the outcomes of the evaluation and then formulate our objectives. So, the evaluation could not have come at a better moment.'* (P1) A policymaker from another province confirms this: *'During the evaluation, we were writing our Nature Vision. We chose the same issues: nature and society, and nature and economy. So it matches. In our final version we have integrated the recommendations of the evaluation. Therefore, the evaluation was a valuable addition to our process.'* (P4)

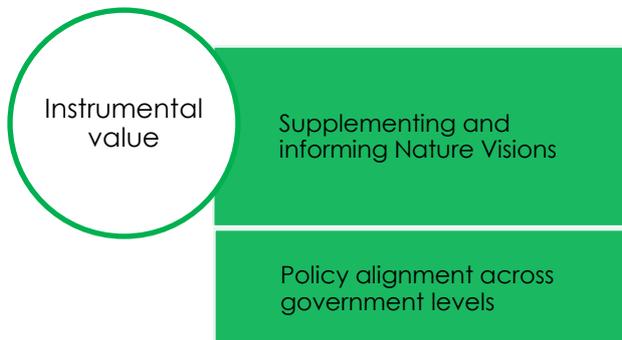
Policy alignment across government levels

The evaluation also made an input to the collective process of nature policy. As one of the commissioners from the Ministry says: *'Next, the provinces and us will discuss how we can translate*



Scottish highland cattle at the Oostvaardersplassen of province Flevoland

the recommendations into policy. The evaluation helps structure these discussions.' (C2) In these ways, the evaluation directly contributes to the process of policy development and the alignment of policy processes across different levels of government.



5.1.5 Strategic value of the evaluation

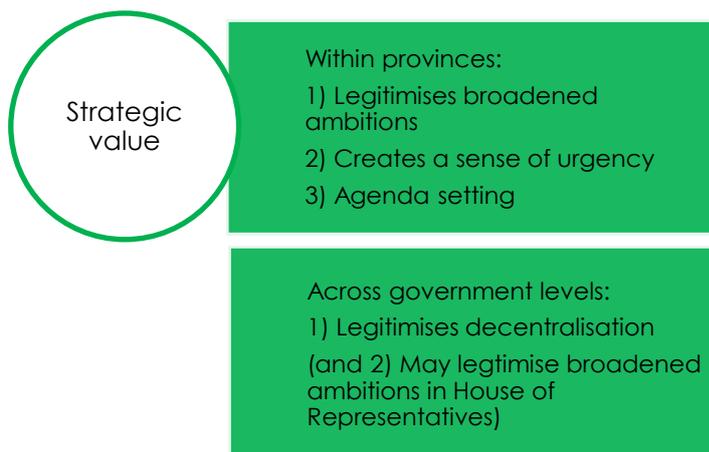
Within provinces

The strategic use of the evaluation lies in its ability to justify choices that have already been made or create a sense of urgency and set the agenda for new directions, especially in communicating with colleagues, provincial administration and societal partners: *'It is very supportive that these broader ambitions of the Natuurpact – nature and society and nature and economy – are now much more in the limelight. That is very helpful, also internally. A lot still needs to be done in these areas, so the evaluation really lends support to discussions about this.'* (P1) This applies not just to immediate colleagues, but also in discussing nature with policy professionals at higher administrative levels. Furthermore, the evaluation also has strategic value in collaborations with societal partners: *'In talking with societal partners it is highly useful to be able to say: "PBL wrote these recommendations, so..." It, in a sense, legitimises some of our work.'* (P3)

Across government levels

A representative from the commissioning parties confirms that this urgency is felt at multiple levels of government and that the evaluation helps to set the agenda in this regard: *'The evaluation really pinpoints areas that still need a lot of work. Connecting nature and society and nature and economy for instance. Even though we already knew that, we still need more effort there, if it is written down in an evaluation, and sharply formulated, it creates more urgency to really do something with it.'* (C4)

Moreover, though the evaluation makes no judgement on whether it is better for nature policy to be centralised or decentralised (which was explicitly not part of the evaluation), involved actors from different government departments and administrative levels perceive the evaluation as having legitimised the decentralisation. This strategic value extends to higher administrative levels, up to the House of Representatives, where the 'success' of the decentralisation is a topic of heated discussion. This success translates mainly into progress on biodiversity levels, not to the other societal ambitions. Deputies comment on how appropriation of this broadened perspective of nature policy by the House of Representatives depends on whether progress is made in biodiversity levels. Thus, the evaluation has strategic value at high administrative levels in legitimising decentralisation, and may eventually also legitimise the broadened ambitions of Dutch nature policy at these same levels.



5.2 Remaining informational needs

As discussed earlier, based on our theoretical framework we expected the evaluation would generate predominantly conceptual and instrumental value, to realise knowledge-enriched policy practice. We hypothesised this would be realised through alignment between the informational needs of policy professional and the content of the evaluation research. Therefore, besides analysing the participants' reflections on the conceptual and instrumental value, we also analysed their remaining informational needs through inventorying their learning questions. These learning questions concern topics that were repeatedly raised by participants and are therefore understood as persistent questions or tough issues. These tough issues were not within the scope of the Natuurpact evaluation and represent informational needs that so far remain unaddressed in the context of the evaluation. Such learning questions and tough issues may become compromising at a later stage. We observed three themes of tough issues articulated by the participants. These relate to the new broadened discourse on the ambitions of nature policy, static institutionalised biodiversity targets, and the substantiation of new roles.

5.2.1 Institutionalising the broadened ambitions

Institutionalised policy goals

Provinces commented on the absence of institutionalised (and thereby shared) policy goals regarding the two new ambitions. Without these goals, no-one can measure whether progress is being made in attaining the ambitions. Furthermore, to measure progress appropriate indicators need to be developed. There is concern for using indicators that are too rigid, and thereby not appropriate for monitoring societal engagement and an improved relation between nature and economy. There remains an informational need regarding how to develop relevant, non-rigid indicators to monitor progress of the two new ambitions. However, prior to institutionalising the broadened ambitions (and its indicators), these need to be appropriated by all relevant stakeholders.

Appropriation of new discourse by other stakeholders

In paragraph 5.1.4 we discussed how the broadened ambition of nature policy has become more shared by members of the Workgroup. For many years, biodiversity has been the principal focus of

nature policy and the formulation of two other ambitions raised many questions: what is the ultimate goal of nature policy, and what do we understand as tools to get there? What exact purpose does increasing societal involvement serve? The value of greater societal (including economic) engagement in support of biodiversity is not a new idea. One perspective holds that an improved relationship between the economy and nature and greater societal involvement increases societal support for nature, and so ultimately has a positive effect on biodiversity. In this perspective, the 'new' ambitions are considered as a means to obtain the biodiversity end. A different perspective holds that all three ambitions are equally important. Developing business models, monetising nature and increasing society's involvement are regarded as ends in themselves, even if this implies a (short-term) trade-off regarding biodiversity. This latter perspective has been largely inconceivable for nature policy in the past century.

Our analysis of the learning questions shows that, especially during the initial phase of the evaluation and drafting the evaluation plan, the first perspective on the ambitions prevailed. Some policymakers seemed uncomfortable with the new ambitions and were sceptical of their value for biodiversity: *'What do new societal initiatives really contribute to biodiversity?'* (P20) Interestingly, after the first evaluation phase we observed fewer comments questioning the broadened ambitions as ends in themselves. Corresponding with the reflection of the Workgroup, among policymakers involved with the evaluation the new discourse has become more shared.

The members of the Workgroup indicated that the shared ambition is limited to those who have been more actively involved with the evaluation. In our analysis of learning questions, we frequently saw questions signalling difficulties originating from different perspectives on the goals of nature policy and their interrelations. A provincial policymaker reflected on difficulties experienced with colleagues who have been less involved: *'To them, there remains tension: "What are we doing this for? Does this contribute to nature?"'* (P2) Furthermore, despite some indications the new discourse is appropriated at higher administrative levels, in some provinces the administration seems to adhere more to the prevailing perspective, thus hampering provincial policymakers in taking action towards the new discourse. Similarly, the analysis shows policymakers sometimes find it hard to put the broadened perspective into practice as societal partners generally also accept the prevailing policy discourse. Learning questions on how to deal with these conflicting perspectives on policy discourse remained posed throughout the evaluation.

5.2.2 Emphasis on static biodiversity targets

Second, we observed learning questions relating to the internationally agreed upon nature norms. Specifically, how their static and inflexible character have only a limited match with nature's inherent dynamics. In addition, they sometimes obstruct the realisation of societal engagement and their focus on demands means that nature outside the designated areas tends to be overlooked.

During the evaluation, some policymakers carefully questioned whether the VHR targets are too static in relation to natural ecological processes: *'Do the international static goals really fit with our dynamic nature?'* (P20) This provincial policymaker explains how preserving a specific habitat-type in a designated area while ecological processes (also, for instance, climate change) are causing the type to naturally 'move', is money wasted. The inflexibility with which these targets are maintained is also perceived as inimical to social engagement: *'If we can spare a farmer or make concessions to a local*

initiative by relocating a designated area... I mean, if it doesn't really matter in biodiversity... but the system scarcely allows it.' (P21) And a commissioner adds: *'In the past three years you see that the provinces put the preservation-goals more on the table... They seem to realise that nature is about more than the preservation-goals and the hectares. [...] If such targets stand in the way of serving the larger purpose – improved nature – then this should be open for discussion.'* (C1) There are thus visible developments in how these targets are perceived. Furthermore, the VHR targets are taken as a benchmark against which the potential of provincial policy plans was measured in the evaluation. As EZ is accountable to the EC for obtaining these goals, the VHR goals would obviously play a significant role in the evaluation. The provinces focus their efforts mainly on attaining these obligatory goals, since government financial support is exclusively for realising European biodiversity goals. The weight given to these goals is sometimes questioned, specifically in terms of its effect on biodiversity levels in areas outside the Natura2000 and the Dutch Nature Network. Learning questions or concerns expressed are, for example: *'If we focus only on realising the static goals, we might lose sight of other biodiversity.'* (P20) and *'How does emphasising the VHR affect other nature? Are we neglecting it?'* (P22) The final report also concludes that this focus constrains the time and resources available to invest in policy innovations.

The imposed nature norms derived from the EU and their prominence in the evaluation scope may thus have some undesirable consequences. While there were recurrent learning questions on these norms, the issue was not much discussed in the workshops or interviews, and questions on the matter occurred sporadically with little follow-up. Though this theme fell outside the scope of the evaluation, the limited discussion on the matter may also be explained from a New Public Management (NPM) perspective: explicit standards support output control, facilitating the transfer of such targets to lower (government) levels (Hood, 1995). The provinces obtain funding from the national government in order to achieve European biodiversity targets and, in collaboration with societal partners, the targets help the provinces in making clear-cut agreements. Pragmatically, putting these targets (e.g. the size and location of specific areas) up for discussion may compromise existing and future arrangements between provinces and public and private partners, and puts the provinces at risk of having their funding reduced if they fail to reach the targets. In addition, they perceive achieving the targets as obligatory *'homework'* (R6) from national government in order to guarantee that the decentralisation of nature policy will not be reversed. It seems that policymakers are aware of the limitations of the existing targets, but there is little motive or urgency to challenge the status quo.

A provincial policymaker confirms there are actors who prefer to maintain current practices: *'It's just rather set in stone in some provinces: occupations are built on this – for some organisations, their entire position is based on the number of hectares they own for which they are reimbursed. It's not in their interest to defer from this.'* (P6) This illustrates how deeply this structure is embedded. It was added, however, that there is discussion on this issue, though not in the context of the Natuurpact evaluation but relating to the annual provincial Nature Progress Report (in Dutch: Voortgangsrapportage Natuur, VRN) implemented by the IPO. The VRN reports on provincial nature progress based on a range of ecological indicators. Interestingly, it has been agreed that for 2017 the focus not only on quantity (such as nature hectares), but also more on the nature quality (such as species, ecosystems and environmental conditions) than on quantity, which suggests some progress in the discussion. Some developments on this informational need are thus visible, though outside the scope of this evaluation.

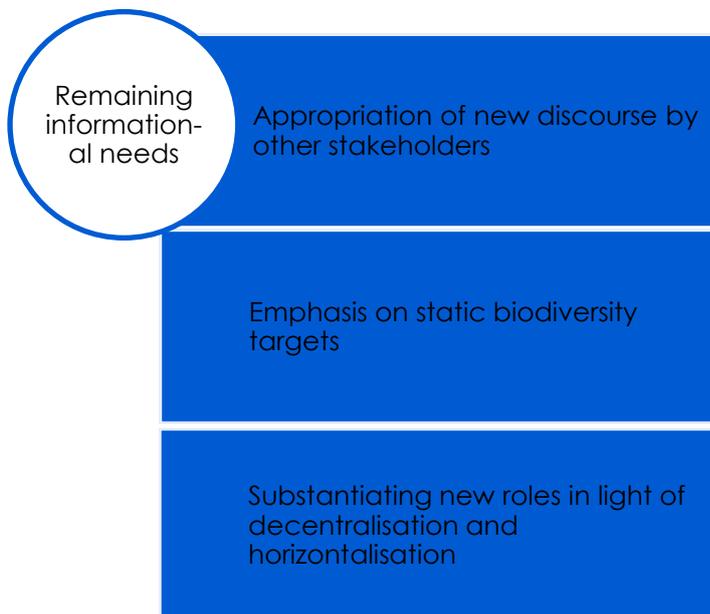
5.2.3 Substantiating new roles in light of decentralisation and horizontalisation

A third theme of tough issues we observed concerns the new roles of the provinces and national government following decentralisation. Learning questions on this issue related to shaping new partnerships between provinces and (new) societal partners. As a consequence of the decentralisation (i.e. increased multi-level government) but also horizontalisation (i.e. increased multi-level governance character) of nature policy, provinces are required to take on a new role. They now function more as a director or facilitating government body, rather than an implementing or executing one. They are thus required to engage in multi-stakeholder processes. At the same time, they are legally obliged to invest in attaining biodiversity goals and are concerned that in taking on a more facilitating role, they will have less control over policy outcomes. For instance, during a personal interview by PBL researchers with a provincial policymaker (P7) on success and failure factors in policy strategies, he reflected: *'How do we offer the required space to civic initiatives in nature conservation areas if these initiatives might have a detrimental effect on the preservation-targets in that area? Are we... 'allowed' to reposition those targets, or let them go altogether?'* The term 'allowed' is meant literally: the lack of judicial clarity further complicates this issue. During the workshop Action perspectives, another policymaker commented (P8): *'Joint-decision making is important, but to what extent? Concerning the international goals: we can't compromise on targets we are held accountable for by national government.'* Provincial policymakers asked the learning questions relating to this issue constantly and throughout the entire evaluation constantly. There seems to be a degree of strategic task certainty (provinces are aware a more facilitating role is called for, which requires finding common ground and shared goals and plans), but simultaneously a level of functional task uncertainty (how should this new role be operationalised?).

Furthermore, decentralisation and horizontalisation also affect the role of national government in the nature policy system. In the Natuurpact it is agreed that national government would remain the 'system responsible' but the learning questions suggest that this role and its properties are very unclear to provincial policymakers. Learning questions, for instance, included: *'What is exactly the responsibility of EZ in this regard? Representatives themselves hardly seem to know...'* (P18) and, more relating to the evaluation, *'How is national government involved with the evaluation – a large sum of the nature tasks is their responsibility.'* (P19) Representatives of national government themselves agree they are unsure of the role of EZ following decentralisation: *'Within our department we also have discussions on, well, nice this "system responsibility" and "director's role", but what does it really entail? [...] A lot has gone to the provinces, and we're searching for what is left, what is appropriate? You can really tell this is really in development still.'* (C2) Another illustrative example is the discussion during an informal administrative meeting, where administrators and deputies deliberated on who is responsible for formulating and concretising goals for the two broadened nature ambitions: national government, or the provinces?

According to both national and provincial policymakers, how the role and responsibilities of national government relate to those of the provinces received too little attention in the Natuurpact evaluation. The final report discusses mainly how national government policy frameworks affect (inhibit) the implementation of new provincial policy strategies. Representatives from national government were relatively little involved during the evaluation because of the uncertainty about its new role. One of them reflects the evaluation might have helped in this regard, but also recognises that other aspects of nature policy took precedence. *'I suppose that would be more suited for the next evaluation, this*

round has been a lot on the substance of nature policy, so the next one may be more on process: relations between the provinces, between us and the provinces, and our role'. (C2)



5.3 Reflection: knowledge-enriched policy practice

At the outset of the reflexive Natuurpact evaluation, it was anticipated that it would contribute to knowledge-enriched policy practice, operationalised by conceptual and instrumental value. In all, our analysis shows indeed these values have been manifested, in more ways than originally anticipated. As we predicted, the instrumental value of the evaluation at this point in time seems limited. The conceptual value is substantial and was esteemed by the participants. Three further types of value are also apparent: network, affective and strategic. Table 5.1 summarises the values that have been manifested and highlights the proliferation of unanticipated values.

Another interesting observation may be made with regard the levels at which the values have meaning. As the Natuurpact evaluation took place in the context of a decentralised policy, it differed from a standard evaluation in the sense that it did not address only one policy or one programme, but rather 12 provincial policies in development and their aggregate (potential) effects. The interaction among the provinces is thus considered very relevant, as it is only through the cumulative effects of the separate policies that it will be possible to achieve the national policy goals. It is therefore equally relevant to consider the value of the evaluation at the collective level, at which relationships are strengthened and ideas are exchanged. Our analysis shows that indeed many of the values mentioned by the participants have also been evident at the collective level. Table 5.1 summarises the different values of the Natuurpact evaluation according to its participants, at both levels.

Table 5.1 Summary of the value-types manifested in the Natuurpact evaluation, presented at provincial and collective level. The values in pink were unanticipated and not documented in the literature on reflexive evaluation.

	Network value	Affective value	Conceptual value	Instrumental value	Strategic value
Provincial level	Network expansion	Relief and reassurance Sense of togetherness	Single-loop: <ul style="list-style-type: none">• Current state• Variety of strategies Double-loop: <ul style="list-style-type: none">• Underlying assumptions• Frame reflection	Supplementation and informing Nature Visions	Legitimisation of broadened ambition Sense of urgency Agenda setting
Collective level	Strengthening of inter-professional relations across government levels	Build trust across government levels	Conceptual coherence Shared ambition	Alignment of policy processes across government levels	Legitimisation of decentralisation May also work to legitimise broadened ambition at higher administrative levels
Relation of manifested values to expected outcomes	Learning Horizontal accountability	Learning	Learning Knowledge-enriched policy practice	Knowledge-enriched policy practice Increased policy impact	Increased policy impact Horizontal accountability

It is important to note that the different types of value are interrelated and affect each other. The participants highly valued the network and affective value of the evaluation; they expanded their own networks, strengthened existing relationships and also found comfort and recognition in mutual struggles. Network value in this regard is conditional for realising affective value. In turn, network and affective value ensured the engagement of participants in interactive learning processes through which the conceptual and instrumental value manifested. The strategic value is the ‘odd one out’ as its manifestation was not dependent on or supported by the other values per se, but was nevertheless much appreciated by provincial policymakers.

5.4 Reflection on theoretical framework in relation to findings

The final row of Table 5.1 shows how the manifested values relate to the expected outcomes of the Natuurpact evaluation (also see Figure 3.1). We observe that all expected outcomes have been more or less attained and have gained meaning during the evaluation. Now, we take the notions and use these to reflect on and improve our original theoretical framework.

We see the concept *knowledge-enriched policy practice* has gained meaning, specifically regarding conceptual value for which both knowledge on the current state of affairs regarding goal attainment, as knowledge on policy strategies to obtain these goals make up an important part. Furthermore, conceptual value also has meaning regarding increased insight in underlying frames that explain individual perspectives and actions. We see that the policy practice related to the Natuurpact

concerns both the practices of individual provinces, and the developing, shared policy practice across government levels. Therefore, we have added a *learning policy network* as an outcome of reflexive evaluation. Both the conceptual, affective and the network value converge in this outcome.

In our original framework we positioned learning as separate outcome. With the new conceptualisation of knowledge-enriched policy practice we described above we hold positioning learning as a separate outcome is a pointless addition. Instrumental value directly relates to *increased policy impact*, which is – at this point in time – outside of our analytical reach.

Furthermore, in our original framework we describe horizontal accountability to emphasise reflexive evaluation aspires not only to provide insight for accountability purposes towards commissioners and sponsors, but also towards all other actors who may experience consequences of the executed policy. We expect further substantiation of the ambition regarding increased societal engagement will give horizontal accountability more prominence in the following evaluation period. We have observed policymakers expect the current evaluation shall have strategic value in terms of generating political and societal support for made policy decisions. Therefore, we replace horizontal accountability with *political and societal support*. Figure 5.1 shows the adapted framework of the outcomes of reflexive evaluation including the positioning of the observed values in relation to the newly conceptualised outcomes.

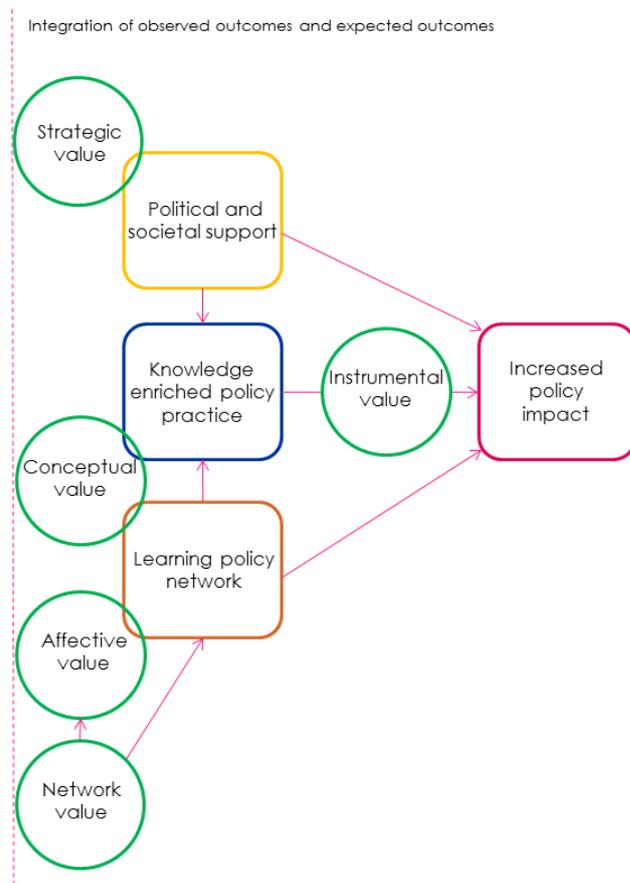


Figure 5.1. Adapted conceptualisation of the outcomes of reflexive evaluation (also see right-hand side of Figure 3.1), integrating the observed outcomes (green circles) and the expected outcomes after Van Veen et al. (2016).

6. Observed factors that contributed to value

As we explained in chapter 1, reflexive evaluation was a new endeavour for PBL. For this reason, the authors were commissioned to develop a theoretical framework on reflexive evaluation to support the researchers in shaping the evaluation design and ensuring the evaluation achieved its potential. In this framework, we proposed several factors as conditional for the success of a reflexive evaluation and the researchers used this information to guide the evaluation process.

We discussed factors such as the selection of stakeholders for participation, followed by stimulating their engagement and participation in the evaluation. Finally, we discussed realising continued alignment between evaluation research and policy practice as important factor. In this chapter, we follow this same line and first discuss who were involved, how stakeholders were engaged to participate, and how the evaluation research was aligned with policy practice. This final category relates directly to the manifested values discussed in chapter 5, and is structured according to those values. In the following texts, we distinguish between factors the evaluation researchers actively steered upon, and unanticipated factors originating from the contexts in which the evaluation transpired. We conclude this chapter with a reflection on the theoretical framework in relation to practice.

6.1 Stakeholder selection

6.1.1 Intended end-users: the provinces

An important characteristic of reflexive evaluation is its inclusive nature: it aspires to accommodate a diversity of stakeholders and perspectives. In our theoretical framework, we state that the stakeholders who should be involved should for the most part be the primary intended end-users of the evaluation findings. The researchers identified these as the provinces. For this reason, the role of national government (EZ) and societal partners was limited this evaluation period. In the following sections, we first discuss the role of EZ and societal partners in this evaluation, and then turn our focus to the provinces.

6.1.2 Other relevant stakeholders

EZ

Naturally, EZ had a role in the evaluation since it commissioned it. In this regard, representatives of EZ (together with the IPO) guarded the initial assignment and scope of the evaluation. However, EZ was not just a commissioner but is also what has been termed the 'system responsible'. Decentralisation means that its role and responsibilities have also changed. Representatives were, however, mostly involved as observers, to keep on top of the evaluation processes. The history of distrust between EZ and the provinces led EZ to keep a low profile: *'With the relations as they are, I think we felt the need to give the provinces the space they needed to take on their new roles.'* (C2) As paragraph 5.1 also demonstrates, this meant that EZ has not become part of the 'Natuurpact community' that developed

among members of the provinces. Furthermore, representatives from EZ indicated that the evaluation did not meet their informational needs to any great extent.

Societal partners

Societal partners were also perceived as important players and initially the intention was to engage them more actively throughout the evaluation process. However, their role in shaping the evaluation was limited and more of an informative nature. The reasons for this were the sensitive relations between the provinces and commissioners (and PBL, by extension). The involvement of societal partners in all evaluation events was not always deemed appropriate as it might have disrupted the soft space in which sensitive information could be discussed: *'With what already was happening, I think it was already a lot. If we would have given societal partners a larger role, it would have affected the levels of trust, it just would have been too much.'* (R3)

Societal partners were interviewed during the preliminary evaluation phase to gain a sense of their position on the Natuurpact agreements and the evaluation and they were involved in the first phase of the evaluation. Interestingly, the initial plan was to conduct multiple separate learning sessions with provinces and societal partners, prior to conducting a collective session dedicated to jointly determining research questions. However, the commissioners were not willing to invest the required extra time as they held low expectations regarding the added value. They were concerned the needs of the provinces and the societal partners would diverge too much, resulting in an exceedingly complicated evaluation assignment. They preferred to focus on making quick progress, and decided to drop this plan for now: *'The commissioners didn't want it, they didn't see the point at this moment in time. They were concerned for a lot of extra complexity, costing too much time.'* (R2) Instead, societal partners joined the provinces in the second and third learning sessions. Furthermore, the societal partners were involved in data collection (during case studies on innovative policy strategies, and interviews on experiences with policy implementation). They were, however, excluded from the group-review sessions, which were attended only by provincial policymakers. In these sessions, the preliminary results on the *ex-ante* assessment were shared, containing sensitive and potentially compromising information on provinces – only at two sessions was a representative from EZ in attendance. The researchers felt they had little choice regarding this matter; more involvement of societal partners might have disrupted the participation of the provinces, and was at times actively hindered by commissioners and provinces.

Societal partners were invited to the workshops Innovative Policy Strategies and Action Perspectives. Interestingly, not all parties attended; not all of them seem to perceive urgency for participating with the workshops. Policy professionals noticed their absence: *'It feels somewhat useless to hold a discussion on water, while there's no Water Board present.'* (P20) This was also noted on agricultural representatives during a session on nature-inclusive agriculture: *'Too often we [policymakers] discuss their [farmers, other agricultural stakeholders] practice without them present. Then they'll say we've decided stuff behind their backs again and excluded them on purpose, and then we're the ones to blame again. These conversations should be held collectively.'* (P18)

The limited role of societal partners in the evaluation so far may have repercussions in a later phase. As paragraph 5.2 demonstrates, developments in policy discourse seem limited to those more involved with the evaluation. Researchers observed that some societal partners have expressed their

discontent and ‘feeling left out’. Not engaging other important players in a process of knowledge co-production may result in obstructive behaviour in later phases as these actors have not been part of newly developed bodies of thought.

From the perspective of the provincial policymakers, diversity in participating stakeholders is also valued: *‘Diversity is a good thing. What is happening with those parties, water boards, big private organisations... What is their position within this evaluation? Are they learning too? It’s difficult because there are so many parties, but taking the two new ambitions into account, it’s only becoming more relevant to involve them.’* (P5) Better engaging societal partners in the continuation of the evaluation is important to ensure that it is also aligned to their perspectives and promotes mutual learning between provinces and societal partners.

Table 6.1 Factors for engaging other stakeholders

Engaging EZ and societal partners	Factors steered by researchers	Contextual factors
Contributing factors	<ul style="list-style-type: none"> • EZ involved in all evaluation phases, largely took on an observer role • Societal partners involved during evaluation, albeit primarily in informative role • Societal partners asked to give formal approval of evaluation plan 	<ul style="list-style-type: none"> • EZ was commissioner of evaluation, and therefore involved ex officio
Hampering factors	-	<ul style="list-style-type: none"> • Involvement of societal partners not deemed appropriate by commissioners and provinces, due to sensitive information • History of distrust compelled EZ to keep a low profile

6.2 Stakeholder engagement

6.2.1 Identifying the ‘true commissioners’

From the outset, the evaluation was intended to encourage learning by the provinces, which are the primary intended end-users of its findings. To engage them in participation, the researchers realised the need to encourage their ownership of the evaluation. However, as discussed in chapter 2, since the provinces do not perceive the IPO as their representative, their distance from the evaluation was initially large. Though EZ and IPO are the formal commissioners of the evaluation, the researchers realised that for the provinces to become engaged, the provinces themselves ought to have a commissioning role. As one of the researchers says: *‘The realisation that not the IPO was a just representation of the provinces, but that we really needed the provinces themselves to join our table was such an eye-opener. That made a huge difference.’* (R3) Consequently, the evaluation researchers

informally appointed the provinces as 12 individual commissioners, with a say in the scope and design of the reflexive evaluation. The evaluation researchers perceived this and the subsequent actions as crucial for promoting provincial ownership of the evaluation: *‘Without it, we would have conducted an evaluation with just some interactive aspects... Likely, the intended learning-part of the evaluation would have stranded, and it would have slowly but surely fallen back into a more traditional impact assessment.’*(R3) By giving the provinces a role in the evaluation process, the researchers encouraged ownership and active participation.

6.2.2. Involvement of provinces during different phases of the evaluation

In the literature on reflexive evaluation (and comparable evaluation approaches) one of the proposed premises is that the evaluation is designed and conducted collaboratively by evaluators and intended end-users. This implies they are involved and have an equal say in all evaluation decisions. We asked provincial policymakers and evaluation researchers to indicate how they perceived the level of involvement of the provinces in the evaluation phase. Their perspectives largely overlapped and are summarised in Table 6.2, which shows hypothetical levels of involvement ranging from more traditional evaluation approaches to a reflexive approach.

Table 6.2 The level of involvement of the provinces by evaluation phase (in green), as perceived by the provinces and evaluation researchers and indicated in green, reflected against the hypothetical levels of involvement ranging from low (traditional evaluation) to high (reflexive evaluation).

Level of involvement of participants →	Low (traditional evaluation approaches)		High (reflexive evaluation)
Research phase ↓			
Phase 1: Determine evaluation objectives and methods	Determined by researchers (methods)	Participants are consulted (evaluation objectives)	Jointly determined
Phase 2: Data collection	Participants are not involved, passively provide data	Participants are consulted, actively provide data	Data is jointly collected
Phase 3: Data analysis and interpretation of the findings	Receive results	Give feedback on results	Give feedback, shared interpretation and draw shared conclusions
Phase 4: Dissemination of the evaluation findings	Receive report (researchers are sender)	Vision of participants is visible (e.g. case stories), researchers are sender	Jointly written report (both are senders)

Phase 1: Determine evaluation objectives and methods

The provinces (and societal partners and EZ) were involved during the first evaluation phase in determining the evaluation objectives and its approach. The first three Learning Sessions were designed to help them articulate learning questions and jointly formulate the main evaluation research questions. In practice this turned out to be no easy task; researchers found it difficult to directly translate the participants’ input into feasible research questions. They commented that an extra session would have been needed to make this a fully collaborative process, but the commissioning parties decided against this. Nevertheless, to address the provinces’ informational

needs, the researchers used their preliminary research on provincial policy plans and interviews with Deputies and societal partners to further inform the evaluation scope and demarcations. The Workgroup Nature Policy formally approved the resulting evaluation plan.

Regarding the evaluation approach and its methods, there was limited deliberation. Most researchers felt that selecting the research approach was their scientific prerogative. As a researcher explains: *'They unlikely have any expertise in that field, why then give them power to decide? We all see the importance of co-creation, but that would just impair the quality of the evaluation.'* (R3) Moreover, limiting the influence of the participants in deciding on research methods was important to the researchers to safeguard their scientific independence: *'We can't let them determine the methods... It would be too much as if we're letting them decide for themselves on which basis they would like to be evaluated. We're an independent assessment agency; some things should be left to our expertise'* as a researcher (R4) illustrates. Interestingly, though most participants agree that the final say on methods lies with the researchers as *'...it is PBL's evaluation.'* (P17) other provinces said that the evaluation would have benefited if they had been consulted on these decisions. For instance, a provincial policymaker explains: *'We had no say in the MetaNatuurplanner and for that model I do think, well guys... It's all very much on a national scale. And during the groups-review it also showed that a lot of things the provinces undertake and think are important, don't show in this model...'* (P1) More discussion on the research methods might have further contributed to the conceptual and instrumental value of the evaluation (to which we return later), or might at least have contributed to better management of expectations in this regard.

Phase 2: Data collection

Initially, it was not planned to collect data together with the provinces. However, as requested data were not provided (due to lack of trust and understanding, as discussed in paragraph 6.1.2), the provinces decided we should visit each of them individually. During these bilateral consultations, the provinces and researchers negotiated the correct interpretation of provincial policy plans. This enabled the provinces to feel included during this evaluation phase.



'Tools' to help formulate action perspectives during collective workshop, during phase 3

Phase 3: Data analysis and interpretation of the findings

This evaluation phase was experienced as the ‘most participative’ by policymakers and researchers alike. Respondents mostly referred to the group-review sessions, which were greatly valued. By presenting the preliminary findings and being open and transparent about the analysis, evaluators helped the provinces feel that they still had a say in the research and felt ownership of the process. Furthermore, the workshop Action perspectives and the smaller expert sessions that followed, allowed for jointly formulating action perspectives based on the evaluation findings. These sessions seem to have contributed to shared ideas on actions to take nature policy still further.

Phase 4: Dissemination of the evaluation findings

The final phase seems to have been the least participative. Interestingly, provinces and researchers alike emphasised the importance of limited participation in this phase: in order for the report to have strategic value, its independent status should not be questionable in any sense. We return to this matter when we discuss factors contributing to strategic value in paragraph 6.4.

6.2.3 Organising administrative commitment at multiple levels

Subsequently, based on previous experiences and inspired by public administration researchers, the evaluators recognised that to compel the provinces to participate, there was a need for administrative commitment at multiple levels. To establish this, the researchers sought out ambassadors or agents of the reflexive evaluation, starting at the highest administrative level of the provinces and working towards the management and executive levels. Ambassadors are individuals who convey the importance of learning from the process of policy implementation and help to guard the related evaluation research.

Starting in the preliminary research, evaluation researchers interviewed several Deputies to take into account their perspectives in the evaluation, rather than narrowing their focus on policy executives, in order to promote support at higher levels: *‘Provincial ownership over the evaluation was necessary, on all levels: administrative, management and executive. We knew we had to organise this on the highest level first, and that the rest would then likely follow.’* (R2) Another pivotal action was the meeting organised with Deputy van Dijk (province Gelderland), chair of the Administrative Advisory Group ‘Vitaal Platteland’ (Dutch: BACVP), during which the evaluators gained his support for reflexive evaluation. A researcher explains: *‘We needed someone higher up to give their support, to give weight to the learning-part of the evaluation. So I started making little unions, coalitions so to speak. For instance with Deputy van Dijk. He saw the necessity of a reflexive approach, as well as its potential value. Subsequently, he played a role in appointing a chairman to the Workgroup Nature Policy who also saw the importance of learning, another crucial act for the evaluation.’* (R2) This chairman became another agent in promoting commitment. He is part of three influential groups: the Commissioners’ meeting, the Official Advisory Group ‘Vitaal Platteland’ (Dutch: AACVP, which informs the BACVP) and, finally, he chairs the Workgroup. *‘He ensured alignment between the commissioners and the provinces, bolstering the value of learning to them.’* (R2)

Throughout the evaluation, similar actions were taken to display the weight given to the reflexive evaluation by the top administration, which stimulated active participation at lower levels. For instance, at each workshop the evaluators meticulously thought about *who* should open or close it – preferably someone with status, whose commitment to the reflexive evaluation would inspire

broader commitment. Second, the evaluators underlined the scientific credibility of the reflexive evaluation approach by involving experts from the VU University. The notion that ‘experts on learning’ were involved in shaping the evaluation process and the specific workshops gave more prominence and weight to learning during the evaluation process. Overall, these actions on organising administrative commitment contributed to a sense of urgency and urged participants to join in the evaluation events.

6.2.4 Promoting provincial ownership

The steps taken by the researchers to engage the provinces and encourage their participation were continued throughout the evaluation. For instance, promoting ownership by assigning a role to the provinces proved an effective strategy. A typical example is how the evaluators appointed two members of the Workgroup Nature Policy to share responsibility for aligning the evaluation research to provincial policy practice. A member of the Workgroup had commented on the selection of innovative policy strategies for case studies, stressing that these were not sufficiently ‘innovative’ and therefore of no informational interest to their practice. Rather than processing such feedback in isolation, the researchers made two members of the Workgroup share ownership of this issue and frequently discussed alignment of the evaluation research with them. *‘We knew it was pivotal they held some responsibility as well, to evoke collaboration. By making them a shared owner of ensuring our research is aligned to their needs, they just had to become actively involved.’* (R2)

6.2.5 Building trust

Regardless of ownership and administrative commitment, the researchers were aware that without trust the provinces would pay only lip-service to participation, without genuine commitment. A lack of trust – believing that the other party will refrain from behaving opportunistically – may be a reason for parties to withhold (sensitive information). Visiting the provinces personally during the preliminary study in early 2014 and emphasising that PBL (and WUR) were conducting a reflexive evaluation to meet *their* needs, rather than EZ and IPO’s, was an important first step in building trust with the provinces and furthered other strategies to maintain trust, such as working ‘surprise free’: *‘We understood their context and that, to win all parties over, we had to earn their trust. So we introduced “surprise-free working”, meaning that all decisions and all evaluation findings would be discussed with the commissioners and participants before these would reach the outside world. Specifically findings that may be politically sensitive. This was received well.’* (R2)

The importance of trust remained evident throughout the evaluation. As touched upon in chapter 2, the provinces perceived the PBL as an extension of EZ: *‘They were keen to share their policy plans, they are proud of them. But when discussing how exactly they were planning to implement these plans, also financially, they become more reticent and suspicious: “Why do you want to know that?”’* As such, distrust affected different evaluation phases, for instance during data collection for the *ex-ante* assessment of provincial policy. Some provinces were more inclined to share their plans, while others remained reticent. Consequently, the researchers decided to visit each province and bilaterally collect the data. Some provinces were highly reluctant to share their information, as a policymaker explains: *‘They asked for a lot, also quite detailed financial planning, and for me it wasn’t even clear why they were there in the first place – what do you need all that information for, how are you going to use it? First explain your purpose, then we’ll consider sharing our information with you.’* (P10)

Distrust, also partly due to not being entirely clear about the intended purpose, was a reason why a large part of the evaluation on the cost-efficiency of policy plans was suspended.

6.2.6 Being transparent

To build trust, the researchers aimed to be transparent and open in their research and provided the provinces a 'look behind the scenes' in their work. Sustained transparency on the motivations behind their actions, and how they took into account feedback and comments from the provinces, further contributed. A researcher explains: *'All interaction moments are crucial for trust. As you know, trust arrives on foot, but departs by horseback. We were all very conscious about this.'* (R1)

Being transparent and open was a challenge to some of the researchers. They normally function as a distant, authoritative observer, but were now required to interact with participants on a more equal footing. They placed themselves in a vulnerable position in doing so, and were concerned that the provinces would question and disregard their work, which was found unnerving. In anticipation, the researchers meticulously fine-tuned workshops and practised presentations. We believe that these efforts were critical in engaging the provinces to participate in the evaluation: *'For me that was one of the most important parts of the entire process. Just going through all those questions, collecting all the answers, getting a sense of their analysis – and then being able to discuss them. "Ok, you have these results from all your models, but do we recognize them?" which then would lead to more follow-up questions and more discussion.'* a policymaker (P11) reflects on the researchers' transparency.

The evaluation research was made even more transparent by providing records on the main findings of each workshop and the interviews. Furthermore, the actions taken by the researchers in response to comments from the participants were noted and communicated. However, not all aspects of the evaluation were equally transparent. In particular, the research on policy innovations was less visible to participants, as was the theoretical and methodological substantiation of the research on the potential of policy plans in relation to the biodiversity goals. Perspectives on whether this mattered, however, differed. Some policymakers said that they would have preferred more insight and say in the decisions for the research methods used (e.g. computational models in this specific example), while others regarded this as the researchers' expertise and not their concern. The following two quotes illustrate these different perspectives: *'If it would have been made more clear sooner which models would be used, it would have given us the opportunity to reflect on these models and their suitability. They are more useful on the national level, not so much the provincial...'*(P10), while another policymaker commented: *'I trust the PBL to do a good job, I don't need to look into that. They're the experts.'* (P17)

6.2.7 Researchers' sensitivity

We distinguish one overarching factor that contributed to the others discussed so far: the researchers' sensitivity. Sensitivity goes beyond responsiveness to diverse needs and interests and includes sensitivity to different world views, power relations, framings and conflicts (Sarkki, Heikkinen, & Karjalainen, 2013). Researchers' sensitivity is evident in their understanding of the political arena in which they had to operate, the cultural and organisational differences between provinces, and the power relations that affected the behaviour of both the provinces and the commissioners. This sensitivity helped researchers to anticipate the disposition of stakeholders towards the evaluation –

when there was strategic play, which individuals needed more persuasion, when to push through with plans or when to hold back.

Sensitivity was also evident in the amount of effort researchers put in the in-depth analysis of each province, which the provinces recognised as a crucial factor for the value of the evaluation. One provincial policymaker explains: *‘We have a highly complex financial situation in our province, and they put in so much effort to represent it clearly and adequately. [...] They really wanted to know all the ins and outs, which, for me, increased my faith in the entire process.’* (P9)

Other examples of sensitivity are the extent to which researchers were accessible and willing to provide extra clarification about the evaluation whenever requested. A policymaker reflects on the project leader: *‘...he was very open to our signals and we could call any moment – and he would act on it.’* (P5) Also, on request, the researchers visited several provinces after publication and presentation of the final report to formally present and further support interpretation of the findings. In each province they visited they aligned their story (in terms of focus) to meet the specific demands. A provincial policymaker reflects: *‘I feel they really listened [...] in our province I felt the need for additional clarification of the final findings by PBL itself, which they gladly did as follow-up.’* (P6)

Table 6.3 Factors for engagement of the provinces

Engagement of the provinces	Factors steered by researchers	Contextual factors
Contributing	<ul style="list-style-type: none"> • Identifying the ‘true commissioners’ • Organising administrative commitment at multiple levels • Promoting provincial ownership • Building trust • Transparency • Researchers’ sensitivity 	<ul style="list-style-type: none"> • Concern for the evaluation being used to legitimise recentralisation of nature policy motivated provinces to participate in evaluation events
Hampering	-	<ul style="list-style-type: none"> • Distrust between provinces and commissioners: apprehension in sharing in-depth data

6.2.8 Concern for re-centralisation

As we discuss in chapter 2, the provinces were concerned that the Natuurpact evaluation had been initiated by EZ to monitor provincial progress and to eventually use the evaluation findings to legitimise the re-centralisation of nature policy, including the vertical lines of accountability the provinces fought to have removed. This concern worked both against and in favour of provincial participation: they were reluctant to share in-depth information, but also were motivated to participate in each evaluation event to exert control on how the evaluation progressed: *‘The idea of “Let’s be*

there to ensure nothing goes wrong” certainly was also a motive for a long time’, a policymaker (P11) illustrates.

This is especially interesting as it demonstrates that the concern to be held accountable (with potential repercussions) contributed to active participation in learning processes: accountability can thus also work to enforce learning during an evaluation.

Our findings regarding factors for sustained engagement of the provinces are summarised in Table 6.3.

6.3 Alignment between policy practice and evaluation research

We now return to the manifested value types established in chapter 5. In our theoretical framework we state that in order for the reflexive evaluation to be successful there needs to be optimal alignment between the evaluation research and the informational needs of its participants. To establish alignment, we highlighted several ‘process requirements’, which we refer to as factors. We have demonstrated that the evaluation has manifested different value types. In the following sections we discuss which specific factors contributed to each factor, and what actions the researchers took in this regard.

6.3.1 Factors contributing to network and affective value

Here we discuss network and affective value together since the contributory factors largely overlap: as stated earlier, network value seems at least partly a condition to affective value. The network value of the evaluation was unanticipated; our results show, however, the development of a learning policy network. Factors contributing to this development were the organisation of interaction between stakeholders

Interacting with other stakeholders

For a large share, most value types are attributed to interaction – with colleagues from other provinces and societal partners, but also with the researchers – by participants and researchers alike. Though we discuss interacting with others as a factor for network and affective value, it is also an important determinant of instrumental and conceptual value. As a policymaker reflects: *‘Learning really is triggered by the insights you obtain through interacting with others, you’re working on the same kind of projects and that inspiration and recognition really gave me a number of eye-openers.’* (P9) However, as we also established earlier, engagement of other stakeholders was mostly limited to a number of workshops.

Meeting other stakeholders is thus an important factor for the success of reflexive evaluation. An important premise in this regard is the diversity of stakeholders who participate in it. As stated in paragraph 6.1.1 the diversity of engaged stakeholders was somewhat limited this first evaluation period, especially regarding societal partners and EZ representatives. The turnout of societal partners during collective workshops was not always as high as aspired. This suggests that, should the following evaluation period include a broader perspective on stakeholder diversity, the network and affective value may further increase.



Policymakers in interaction during collective workshop

Table 6.4 Factors for network & affective value

Network value & affective value	Factors steered by researchers	Contextual factors
Contributing	<ul style="list-style-type: none"> • Interaction with other stakeholders • Frequency of interactions • Workshops and group sessions designed for sharing experiences and reflection 	<ul style="list-style-type: none"> • High attendance due to concern for the evaluation being used to legitimise recentralisation of nature policy motivated provinces to participate in evaluation events
Hampering	-	<ul style="list-style-type: none"> • Limited turnout of other stakeholders (such as societal partners) during collective workshops

Workshops and group sessions designed for sharing experiences and reflection

Not just getting together, but also having the opportunity to share stories and experiences of addressing similar challenges, contributed to the networks and affective value. The realisation that other provinces are in a similar phase – in relation to achieving societal engagement – was reassuring, and was able to emerge through discussing difficulties in a soft space. A soft space implies there is a safe environment in which to discuss sensitive information, and where all participants are trusted. This was especially the case for the group-review sessions (attended only by provinces, with an occasional EZ visitor).

The original intention was to visit each province to discuss their preliminary evaluation results. The amount of work that entailed, the pressure of time and the realisation that group sessions would allow far more opportunity to learn from each other, made the researchers decide on four interactive group-review sessions with three provinces at a time. During these sessions, the preliminary results were discussed for each province for validation and enrichment. The sessions also enabled the provinces to interactively reflect on the meaning of the preliminary findings to their individual practices and to exchange experiences on these. Though some researchers felt that not all participants showed ‘...the back of their tongues’ (R6), it appears the information shared was sufficient for establishing a sense of reassurance and togetherness.

Our findings regarding factors for manifested network and affective value are summarised in Table 6.4.

6.3.2 Factors contributing to conceptual and instrumental value

Alignment to informational needs

The intention was to optimise conceptual value by aligning the evaluation scope to the participants’ informational needs by involving the provinces in determining the evaluation scope and demarcations and assessing this alignment throughout the evaluation. As we established in paragraph 6.1.3, jointly determining research questions was not a straightforward task. Nevertheless, the evaluation plan was approved by the provinces and also included a flexible outline that permitted adaptation along the way. Further substantiation in response to emerging issues was permitted, which also further improved alignment as the evaluation progressed. Moreover, the members of the Workgroup chose the discussion topics for the parallel sessions of each workshop, by which alignment of their informational needs was also optimised.



Picture taken during a collective workshop, displaying an ambition of nature policy: meeting international goals

The evaluation researchers made every effort to formulate evaluation research questions that best allowed for collective learning among the provinces. For instance, the researchers sought to make innovative policy strategies a central focal point of the evaluation, as innovations were believed to provide a suitable opportunity to inspire and learn from: ‘They don’t realise it, but using innovations

as case studies is really something new. We really wanted it in the evaluation, because we wanted to inspire learning. It took some persuasion and discussion with the commissioners to get it accepted in the plan.' (R2)

In addition, the researchers combined qualitative and quantitative research to assess the *executorial potential* of the policy plans, taking into account contextual factors that hamper or promote implementing policy along the way, rather than merely focusing on the plans' potential on paper. This was also meant to optimise learning by the policymakers, and to better allow them to anticipate such hampering or promoting factors.

The frequency of interactions

The literature presents the frequency of interaction as essential for building enduring collaborations, both among participants, and also, importantly, with the evaluation researchers (e.g. Mattor et al., 2013). The participants' views on the frequency of interaction were diverse, but overall concur that the number of workshops and research-related activities was high: *'At some point, you just think: "Hmm... another workshop...". It also makes people less focused, because the workshops start to feel ordinary.'* (P5) Simultaneously, they reflect that despite the sometimes initial reluctance to join 'yet another' get-together, the workshops were experienced as energising and inspiring. A participant suggests that perhaps the process could have been '*leaner*' (P11), meaning fewer moments of interaction, but there is little support for also making the process leaner by reducing the scope of the evaluation. Reducing the number of interactions while not also adjusting the evaluation's scope is expected to have a negative effect on the value of the reflexive evaluation. Another participant adds: *'Especially with the new ambitions I think there is now more than ever need for knowledge exchange between the provinces – I don't see how we could do with much less interaction moments.'* (P1)

The researchers themselves felt that the process was labour-intensive and ascribe this in part to the novelty of the reflexive evaluation process; researchers and participants alike had to find their way, get to know each other and each other's practices. Some of the researchers feel the evaluation would have benefited from more frequent interactions to keep participants more engaged with the research progress. An example is the first group-review session. The alignment between what the researchers presented and what the provinces came to hear was not ideal. The provinces found figures and diagrams hard to interpret. Consequently, a large portion of the available time had to be dedicated to reaching a better understanding between the parties. Some researchers believe that for this reason, the evaluation needed an extra moment of interaction with the participants: *'A prior session to explain our models and our approach would probably be better... Now we have to stuff a lot of things in one session, that is a risk.'* (R5) However, another researcher (R4) reflects: *'If we would plan more events, that would compromise our planning – we also don't want to delay our final report.'* Commissioners and provinces saw only limited value in more interactions, which, combined with pressing deadlines, led the idea to be dropped.



Presentation during a collective workshop

A balance was sought between the available time and number of interactions, a familiar consideration in extensive collaborative processes and strict deadlines. Making use of the monthly meeting of Workgroup Nature Policy to further alignment of the evaluation and provincial policy practice was appreciated by participants and researchers. Making more use of such existing structural gatherings of provincial policymakers is a potential way to reduce the number of interaction moments without having to cut out content.

Mutual understanding

Next to aligning the research to the informational needs of the participants, the researchers also ensured the research was relevant to them by investing in correctly interpreting their policy plans. This ensured that their assessment of the provinces was based on data the provinces themselves regarded as correct. This occurred during the bilateral consultations, during which the researchers visited each province to collect data on provincial plans. Though viewed as a rather intense process, the researchers' efforts to correctly interpreting provincial policy plans in detail ensured that their approximation of the provinces in the data analysis was as good as possible, and that findings were recognisable and relevant.

The researchers, who intended to collect data based on provincial policy documents, did not originally plan these bilateral consultations. Having found that the documents were insufficient for the level of detail they required for their analysis, the researchers sent out a call to all provinces requesting the necessary data. The response to this call was exceptionally low: *'The evaluation felt enforced for the provinces, and they perceived it as a way of EZ to monitor their work despite the decentralisation, and hold them accountable for realising the set ambitions or not. They weren't planning to willingly cooperate. It was a message: "Leave us alone", which also showed the idea of a reflexive evaluation – learning from each other – really hadn't landed yet.'* (R4) The commissioning parties eventually decided the researchers should visit each province individually. This proved to be particularly valuable; the consultations allowed the researchers to gather detailed information and further

conceptual alignment, and also provided opportunities for building trust and mutual understanding between the researchers and policymakers.

Bilateral consultations and group-review sessions for in-depth interactive reflection

The bilateral consultations were also an important event regarding double-loop learning at the provincial level. Though the consultations were found somewhat burdensome, provinces agreed they had great value as they encouraged them to reflect deeply on their current plans: *‘You were forced to look back and really reflect... The researchers would ask: “why do you do it this way, did you take this into consideration?” etcetera. And then we’d think “Ai.. Well...”’* (P2) Another policymaker (P9) adds: *‘Those bilateral discussions really gave me the most insight.’* The bilateral consultations thus were important in several regards: for contributing to network and affective value (albeit between the provinces and researchers, not between participants) by building trust and mutual understanding, and for contributing to conceptual value by promoting conceptual alignment and in-depth reflection on policy plans.

It was not only the bilateral consultations that provided opportunity for in-depth reflection, the group-review sessions contributed to this too: *‘I remember the session we had with two other provinces. It was so useful to really dive into each other’s approaches, in such detail.’* (P16) This opportunity to take a step back, reflect and exchange thoughts and experiences with actors other than direct colleagues contributed to an in-depth understanding of policy theory, but also of each other’s perspectives. Specifically, during the workshops there was much attention for discussing personal challenges and sharing experiences. It seems this contributed to exchanges of perspectives and a better understanding of each other’s position with regard to nature policy.

Table 6.5. Factors affecting conceptual and instrumental value

Conceptual and instrumental value	Factors steered by researchers	Contextual factors
Contributing	<ul style="list-style-type: none"> • Alignment to informational needs • Focus on opportunity for learning • Mutual understanding • Opportunity for in-depth reflection and frame reflection (during bilateral consultations and group-review sessions) • Jointly formulating action perspectives 	<ul style="list-style-type: none"> • High attendance due to concern for the evaluation being used to legitimise recentralisation of nature policy motivated provinces to participate in evaluation events • Reluctance for sharing information (during data collection), delaying the evaluation process may be seen as a hampering factor, but resulted in bilateral consultations – which is why we pose it as a contributing one
Hampering	-	-

There were, however, also provincial policymakers who were more critical of these moments. For instance, one saw no added value in such sessions and commented: *'Well, it is always nice to see your colleagues, but it didn't bring us anything new.'* (P10) Another policymaker perceived the evaluation events as learning opportunities, but felt that more explicit attention should have been given to this learning aspect. *'From a transition-discourse point of view, the evaluation would have benefitted from more explicit and purposeful learning moments. We speak of a learning [reflexive] evaluation, so learning on different levels – individual, organisational, systemic – should have been made more explicit during these moments.'* as she points out (P6).

Our findings regarding factors affecting conceptual and instrumental value are summarised in Table 6.5.

6.3.3 Factors contributing to strategic value

Finally, we consider the factors that have contributed to the strategic value of the evaluation.

Perceived credibility of evaluation

None of the participants truly questioned the researchers' credibility. Perspectives on credibility, however, do seem to differ among them. Some policymakers seemed to determine the researchers' credibility by proxy: the authoritative image of the PBL was sufficient for them to trust the researchers to have scientific expertise and to conduct the evaluation independently and with scientific rigour: *'How they exactly do it, I don't know, but it's also not so important to me. I mean, I sufficiently trust PBL to know they'll conduct a proper analysis.'* (P17)

The provincial policymakers agreed that transparency on how the assessment was conducted was an important factor for affirming credibility. As one policymaker (P10) puts it: *'Without transparency I would have disregarded the results immediately, I would've thought "What is this based on and what does it have to do with our practice?"'* Some participants were more critical of the evaluation approach and looked for more detail. For instance, critical questions were asked during the group-review sessions relating to the 'expert judgment' that was made in order to translate qualitative data to quantitative data as input for the computational model. For some participants, understanding exactly how the *ex-ante* assessment was conducted was thus important in order to perceive the findings as trustworthy and rigorous.

Third, the perceived independence of the researchers contributed to the credibility of the evaluation. While working in close interaction with those whose policy is under evaluation, researchers may appear and become biased. As stated above, their independence was not questioned, though it was discussed by some policymakers: *'There is of course a danger of appearing subjective... I feel it has been done with a lot of integrity, but I do sometimes wonder whether we should keep calling it an evaluation instead of guided intervision or something.'* (P12)

The researchers used various strategies to guard their independent status. To start, researchers' roles (assessor, facilitator, mediator) were distributed among members of the evaluation project team, which meant that some were in more direct contact with the participants while others functioned more at a distance (also recommended by Edelenbos & van Buuren (2005)). Furthermore, the

researchers ensured data triangulation by collecting qualitative data not only from provinces (which would allow for bias), but also from societal partners in respective provinces.

That the researchers and the evaluation are perceived as fully independent was highly important to the participants, specifically for the strategic value of the evaluation. To use the findings to legitimise their policy plans or to support certain topics on the policy agenda, it is essential that these be viewed as an independent expert judgment: *‘The evaluation itself is something we can use and say: “Look, it has all these recommendations, great examples, we are making our nature vision, we should use this.” And when it’s written in black and white in a PBL report, that’s a world of difference compared to when some ecologist from the province says it’*, as a policymaker (P13) puts it. Protecting their independence is thus crucial for researchers to produce credible research, and for the participants, for the evaluation to legitimise their work. Some provinces, however, have commented on the final report being *‘too nice’* in this regard; they would have preferred the findings in the report to be framed more urgently, emphasising that there remains a lot to be done with regards to biodiversity levels. A few of the provinces argue this would have given the evaluation more strategic value in the sense of giving urgency to their cause in discussions with colleagues, higher administrative levels and societal partners.

Table 6.6. Factors affecting strategic value

Strategic value	Factors steered by researchers	Contextual factors
Contributing	<ul style="list-style-type: none"> • Perceived credibility of evaluation <ul style="list-style-type: none"> ○ Credible institution ○ Transparency ○ Independence • The right people in the right position 	<ul style="list-style-type: none"> • Status of PBL as authoritative, trustworthy research institute
Hampering	<ul style="list-style-type: none"> • Relatively ‘nice’ framing of findings 	-

6.4 Reflection on theoretical framework in relation to our findings

We observe the concept *end-users* has been interpreted more narrow than generally discussed in literature. The provinces may be perceived as the primary end-users of the evaluation, but in the context of the broadened ambitions of the Natuurpact, for which an increasing number of societal actors - ranging from water boards to agricultural representatives (e.g. LTO), terrain management organisations and NGO’s – is involved a broader interpretation of ‘end-users’ would be in order. The reflexive evaluation of the Natuurpact teaches us, however, this is easier said than done. The second condition, realising stakeholder commitment, and the third, aligning evaluation research and policy practice, turned out to require substantial investment during the evaluation. Additionally, policy practice is not a univocal concept, but rather a policy arrangement of diverse interrelated policy processes across multiple levels and within twelve provinces.

We see the reflexive evaluation has given further substantiation of the second and third condition. While transparency and frequent interaction were also mentioned in our theoretical framework,

researcher sensitivity came less forward in relevant literature, while this concept has proved an important factor for success according to the participants of the evaluation. Furthermore, the evaluation has given insight in the diverse range of interactions to establish alignment and knowledge exchange (ranging from bilateral, group-review and collective learning sessions), and the diverse ways these type of interactions were of value. Finally, we observe that despite scholars advising far-reaching involvement of actors in all phases of the evaluation, in the Natuurpact case the final evaluation reports are owned by PBL and the WUR. The decision not to make the provinces co-owner of the reports has had important strategic value for their practice.

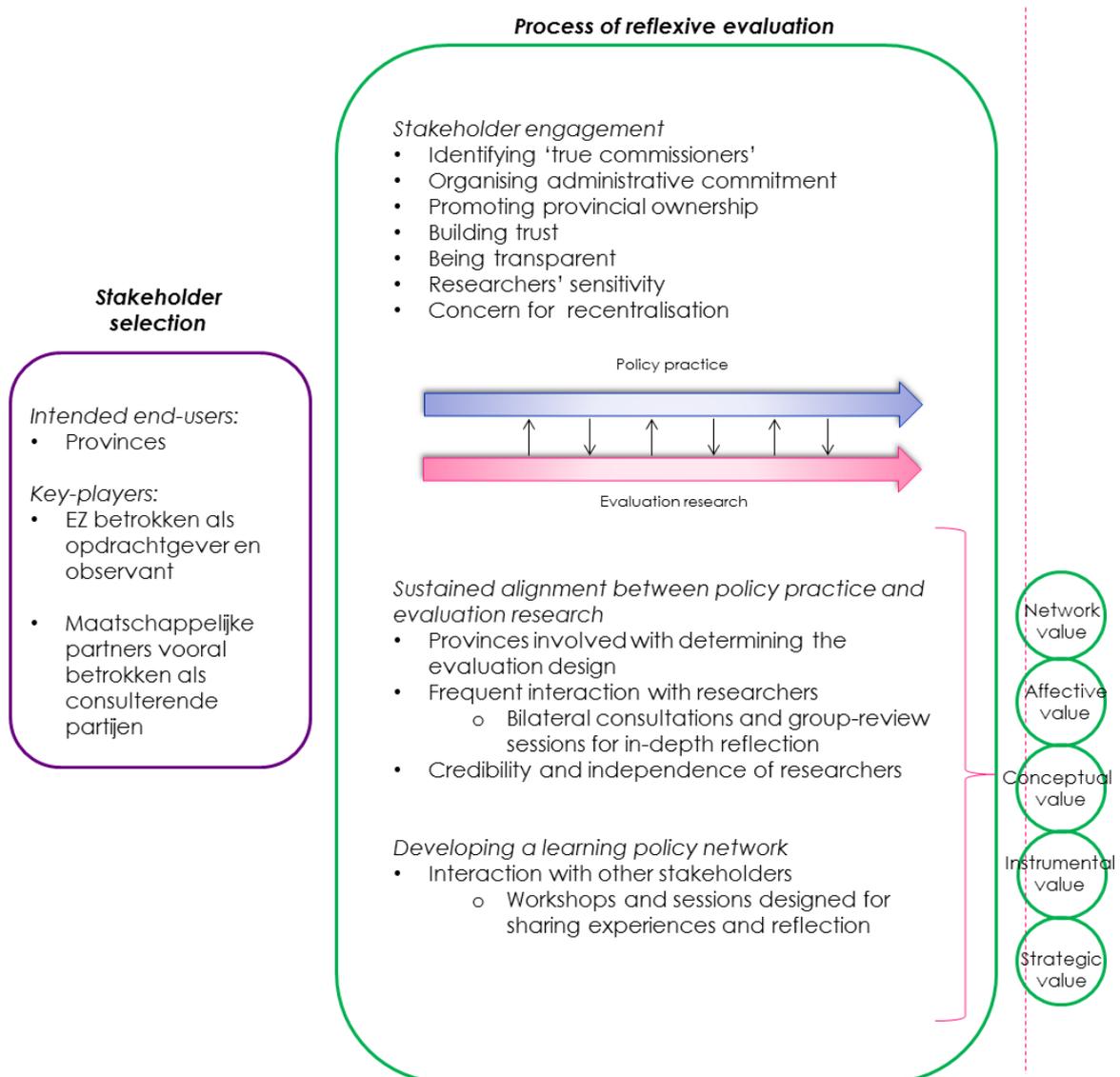


Figure 6.1 The identified factors in relation to the observed values (also see left-hand side of Figure 3.2).

7. Final conclusions and recommendations

This review aimed to obtain insights into the value of the Natuurpact evaluation according to participants and the factors contributing to this value. We were participant–observers during evaluation team meetings, facilitators during multiple learning sessions, had access to documentation of other meetings and exchanges and we spoke to key stakeholders from provinces, societal partners and national government (by means of interviews and focus group discussion). We found that the areas in which its participants found the evaluation most valuable were the conceptual, the affective and the strategic value.

7.1 Overall conclusions

First, participants indicated that they highly valued the knowledge gained and understanding generated as a result of the evaluation, which we understand as conceptual value. This is in particular the case on the individual provincial level; the learning needs of national government (regarding their new role within the newly decentralised nature policy system) are not addressed in the evaluation, which is regarded as a shortcoming both by the provinces and by national government. Nevertheless, conceptual value has also manifested on the collective level, primarily in the form of a more shared broadened ambition of nature policy across governmental levels. We consider the most important contributing factors to conceptual value to be the researchers' sensitivity to the needs, situation and challenges of individual provinces, and their transparency regarding data collection, interpretation and framing. The latter factor was anticipated by the researchers, and they took conscious action to work transparently. Regarding researchers' sensitivity, this factor was less expected to play such a major role (and was also less operationalised) and seemed an almost natural skill of the researchers.

Second, affective value was generated through the evaluation. The decentralisation process is complex, with many unknowns and working routines and cultures that still need to be developed at provincial level. At this level, meeting other professionals who are in the same situation contributed importantly to a more reflexive and open perspective on the decentralisation process and the role of provinces in implementing nature policy. 'Not knowing' became acceptable, rather than uncomfortable, and gave more space to develop and try out different strategies. On the collective level, we perceive signs of increased mutual trust between provincial and national governments. The affective value was unanticipated and seems an outcome characteristic for reflexive evaluation, to which we return later. Crucial factors in this regard were the design of the workshops and group sessions in which the sharing of experiences and challenges was emphasised. The researchers were aware of the value of sharing successes and dilemma's and used this intentionally. Particularly the group-review sessions with groups of three provinces, greatly contributed to a sense of relief and reassurance.

Third, the evaluation clearly had strategic value. On the provincial level, it legitimised the policy decisions relating to the broadened ambitions of nature policy, and gave this change in policy discourse a sense of urgency and a position on the agenda. Though unintentional, informally the

evaluation is perceived by stakeholders to legitimise the decentralisation of nature policy on the collective level and set the agenda for rethinking its goals, introducing a supported move from a mere focus on biodiversity goals to enmeshing biodiversity goals with broader ambitions pertaining to the relationship between nature and society and nature and economy on higher administrative levels (the Dutch Senate and the House of Representatives). A crucial, unanticipated factor in this regard was the perceived authority and independent status of PBL as environmental assessment agency and, by proxy, the independent status of the final evaluation report. Should there have been any doubt regarding this status – for instance, in the case the provinces would have shared authorship over the report, as recommended in literature on collaborative evaluation approaches – the strategic value of the evaluation would have been considerably decreased. The importance of the strategic value, especially on the provincial level, was greater than originally anticipated.

Interestingly, the ‘unspoken’ importance of the evaluation for legitimising the decentralisation process (originating from the tense relation between the provinces and national government) supported the learning processes of the provinces during the evaluation. Whereas accountability and learning are often considered as trade-offs in the literature, here they reinforced each other: accountability (by provinces translated into concern for recentralisation) stressed the importance of the evaluation and led to high attendance and involvement, which in effect created conceptual and affective value. Nevertheless, the concern for recentralisation and distrust between the provinces and national government also affected how openly the parties shared information and spoke about their experiences with each other and with the researchers (who the provinces perceived as an extension of national government). Also, concern for recentralisation resulted in the decision to give societal partners a smaller role in this first evaluation period, as this was expected to further discourage the provinces’ openness. In this regard, accountability and learning hampered one another and likely reduced the conceptual, affective and network value potential of the evaluation.

Finally, we draw attention to the actions taken by the researchers to inspire engagement of the provinces: identifying ambassadors for the evaluation, building trust, their sensitivity to individual provinces’, transparency in their work. Though these factors are not one-on-one relatable to the manifested values, we hold these as paramount for the overall success of this evaluation.



Dutch tulips

7.2 Discussion

Compared to the theoretical framework, we see that network and affective value and values at the collective level seemed especially specific to the reflexive evaluation. Specific to the collective level, values such as the building of trust and strengthening of relationships across government levels have manifested (affective and network value). Moreover, gaining a better insight into other stakeholders' perspectives, or 'frames of reference' (Schön & Rein, 1994), enables policymaking in a pluri-form, multi-stakeholder context, resulting in the alignment of processes of policy development between provinces and across other levels of government (e.g. national, water boards) (conceptual and instrumental value). These unexpected collective values may be characteristic of the particular context of the Natuurpact evaluation, namely a large-scale and complex policy programme (Teisman et al., 2002; Van der Meer & Edelenbos, 2006), in which policy development and implementation take place at different levels of governance in many geographical areas. These values – and specifically the alignment of policy processes across government levels – may therefore be highly relevant and valuable for policy programmes in similar contexts, or in a transition to decentralisation.

Concerning the strategic value of the evaluation, some provinces find the evaluation report too 'nice'. Sharper formulations on the findings would have supported them more in organising directive action in their provinces and thereby would have increased strategic value. However, the researchers have opted for this softer version for several reasons. To start, the phase in which provincial policy transpires played an important role. Provincial nature policy is relatively young; during the start of the evaluation the provinces had been developing and implementing nature policy for four years – relative to the response time of nature to intervention, this is little time. Following, an *ex-ante* evaluation was conducted, focusing on the *potential* of policy plans. This does not allow to draw hard conclusions on policy impact. The tension between provinces and national government (and the perception of PBL as extension of national government) and the concern for recentralisation further made the researchers steer clear of framing that allowed for interpretation of findings in support of recentralisation. This was also important to emphasise the trustworthiness of the researchers and their independence of national government. Finally, the researchers reasoned a negative framing of the evaluation findings would discourage learning processes. The positive framing of the evaluation findings shows similarities with the argument of Van der Knaap & Turksema (2015) on positive policy evaluation. They argue that it is legitimate for evaluation researchers to emphasise the positive policy developments, and 'moving with its participants' rather than focusing on the negatives, especially when an evaluation concerns new policies or changed circumstances in which policy is implemented. Focusing on the positives and successes supports constructive interaction and increases the use of evaluation findings (the conceptual and instrumental value of evaluation) (Van der Knaap & Turksema, 2015). These authors further recommend that a more 'neutral' evaluation should follow, up taking a more critical perspective to policy.

Finally, we observe the values manifested mainly for participants who were more actively involved in the evaluation, such as members of the Workgroup. This resonates with theories on learning such as the concept 'communities of practice' coined by Lave and Wenger (1991). They discuss how a community may emerge in which shared understandings develop, and participation in practice is posed a principle of learning in this regard (Yakhlef, 2010). For the continuation of the evaluation this is a paramount observation: policy professionals and societal partners who are 'left behind' and less

part of these mutual learning processes may in the future oppose decisions made based on the developed shared understanding as they have not engaged in the same cognitive processes.

7.3 Main conclusions and corresponding recommendations

Stakeholder selection and engagement

Societal partners have been involved to a limited extent in the evaluation so far. However, considering the increased complexity of nature policy and the increasing number of actors involved with policy development and execution, it is important to interact with other parties, as they hold relevant knowledge to better understand and grasp the nature policy system. If these parties are not involved, the produced knowledge is expected to be less socially robust, resulting in less ownership over the findings and their use in policy practice and, eventually, suboptimal policy impact.

- For the next evaluation period, we advise to involve a broader selection of stakeholders (such as societal partners involved with the development and implementation of nature policy, but also representatives from national government) in determining the evaluation scope and its demarcations, and the execution of the evaluation research.

We observed meticulous attention for encouraging engagement of the provinces in the evaluation. A number of important factors which contributed to encouraging engagement were identifying ambassadors, creating trust through transparency and researchers' sensitivity.

- When other potential end-users are given a larger role in the second evaluation, new attention is required for encouraging their engagement and creating trust in the evaluators. Careful and effective communication, transparency, identifying ambassadors and organising administrative support will again be important strategies. In the process design of the evaluation we advise to take into account the increased complexity originating from the increased number of involved actors and heterogeneity in perspectives and interests.

Continuous alignment between policy practice and evaluation research

End-users (primarily the provinces) have been more or less involved in all phases of the evaluation research. During phase 1 (determining the evaluation framework and methods) substantial interaction was organised between researchers and the end-users, providing valuable insights, which for the large part have been integrated in the evaluation framework, albeit limitedly systemically. Through frequent interaction between researchers and involved policymakers the evaluation researchers largely aligns to the learning needs of the policymakers, specifically regarding knowledge on plan-potential, executional potential of policy plans and innovative policy strategies. Some of the policymakers felt that the scale on which the evaluation was conducted (national) resulted in suboptimal action perspectives on provincial scale. This is related to the choices for research methods (specifically the Metanatuurplanner), in which policymakers had little say.

- The final evaluation report and this current review offer already a set of potential research questions for the following evaluation framework; it is to be expected that the next evaluation will follow up on whether the assessed potential of policy plans have been realised and how the recommendations of the final report have manifested in policy practice, as well

as to focus on learning with regard to the identified learning needs that are currently unaddressed in the evaluation. On the other hand, we also recommend to carefully conduct phase 1 again, systematically integrating the input of the diverse involved actors. This implies anticipating on the required time for supporting the articulation of learning needs by actors, analysing their input and translating this to shared evaluation research questions..

- Specifically as this next evaluation will likely be accompanied by a larger number, potentially conflicting, perspectives and new learning needs, it is recommended that the process design of the evaluation takes in to account this important translation into the evaluation framework.
- As a consequence of the broadened stakeholder selection, we recommend reviewing the decisions for evaluation research methods, preferably in interaction with involved actors.



Kingfisher

During phase 2 (data collection) and phase 3 (data analysis and interpretation) there was also frequent interaction between researchers and policymakers, contributing to mutual trust and resulting in high quality of information, socially robust knowledge and supported findings. The chosen types of interaction played an important role in this regard. This review demonstrated the value of using diverse types of interactions – such as the bilateral consultations, group-review sessions, collective learning workshops – for different goals. These insights may be used to further inform phase 2 and 3 of the second evaluation. Furthermore, it proved effective and efficient to make use of existing structures, such as the frequent meetings of the workgroup Nature policy.

- Sustained interaction remains important during phase 2 and 3. Diverse types of interactions may be used for different purposes, including the use of existing meeting-structures.

The researchers had the lead during phase 4 (specifically regarding dissemination of evaluation findings); the final report is owned by PBL, and is not a shared publication with the provinces. Policymakers emphasised the importance of this, as it stressed the independence of the researchers, and thereby the credibility of the evaluation findings. In literature, scholars warn for decreased

ownership and support of the evaluation findings as a result of this, but through frequent interaction in the previous phases this was not an issue.

- The strategic value of an independent PBL-report should not be underestimated. We recommend continuation of the used strategies for guarding the researchers independence.
- Simultaneously, the independence of the researches may be at odds with ownership and support of the findings, which originated from a process of interaction and co-creation. Sustained interaction during the previous evaluation phases helps guard against this.

Further developing the learning policy network

An important finding is that the Natuurpact reflexive evaluation has revealed network and affective value; a learning policy network has developed within which actors have shared challenges, struggles and opportunities, supported by mutual trust and reassurance. The provinces and national government have indicated their desire to maintain the learning character of the evaluation. The combination of the urgency of the evaluation (derived from the concern for being held accountable and recentralisation of nature policy) on the one hand, and the opportunity for learning from other players in the field on the other, is what makes the context of the evaluation opportune for facilitating learning on the development and execution of nature policy. During the first evaluation a number of learning needs have been insufficiently addressed, which has affected the realised conceptual and instrumental value of the evaluation. These concern the learning needs of the primary end-users (the provinces), for instance in relation to substantiating the broadened ambition and respective indicators, and regarding the new role as facilitating governmental body. Also, the learning needs of other actors are relevant. For instance, representatives of national government have indicated that they have a learning need on substantiating its role as system responsible. The learning needs of other actors have not been inventoried.

- We advise to continuously align the scope of the evaluation with the emerging learning needs of involved actors, taking into account specific subjects that require learning and which have been unaddressed so far.
- To gain insights in the emerging learning needs, as well as in the learning needs of 'new' actors it is useful to inventory their learning needs.
- Specifically, we draw attention to inventorying and addressing the learning needs of representatives of national government (the ministry of EZ) (as co-commissioner and simultaneous stakeholder within nature policy).
- Monitoring learning agendas of involved actors (and thus their learning needs cq. questions) may improve the insight in the impact of the evaluation on policy practice. Moreover, these insight may inform the third evaluation period (2021-2024). We advise to take up the monitoring of learning needs as explicit evaluation goal. Intermittent reports (bi-annually) on the developed learning agendas supports timely feedback and reflection.
- In the process design of the evaluation, time and space is required to anticipate on taking into account new learning needs; opportunities for realigning planning and budget as a consequence of emerging learning needs may for instance be incorporated in the formal evaluation assignment.
- Should there be decided for a more integral evaluation scope, taking into account all Dutch nature policy instead of demarcating the evaluation to the agreements of the Natuurpact, this likely also gives rise to new research questions and a need for additional expertise.

Maintaining the balance between learning and accountability

The results show that a (fragile) balance has been established between learning and accountability. Diverse learning processes have transpired for the primary involved actors of the evaluation. Simultaneously we observe that the evaluation is used successfully for accountability purposes on several government levels. Learning and accountability both enhance and hamper one another. The tense relations between the provinces and national government, and the perception of some provinces of PBL as an extension of national government, have caused the provinces to translate evaluation for accountability purposes into a concern for retribution and recentralisation. This gave the evaluation a sense of urgency, contributing to the high attendance of provinces during evaluation events and thereby to their learning processes. On the other hand, this concern for recentralisation caused provinces to be less open in sharing information; sometimes they were hesitant in this regard, which complicated and slowed down the evaluation process. The tense relation between the provinces and national government shows improvement. PBL also has gained trust during this evaluation period. Nevertheless, assuming the concern for recentralisation has been resolved would be naive. Especially taking into account the nature of the coming evaluation period, which will be not only of an *ex-ante* character, but also *ex-post*, which will likely increase the accountability-purpose of the evaluation. The *ex-post* character of the second evaluation implies focus will be more on the actual progress provinces have made in relation to the goals of nature policy, which may increase the provinces' concern for retribution and recentralisation. This will have consequences for how learning and accountability relate to one another, and thereby also for the built relation between the researchers and the end-users of the evaluation.

- To maintain balance between learning and accountability we advise to integrate both of these concepts in the process design of the valuation. For example, appointing evaluators who focus on evaluation for accountability of policy (and thereby on scientific rigour and independence) and evaluators who focus on the interaction with involved actors proved an effective strategy that may be sustained the following period.

So far evaluation for accountability mainly concerned the biodiversity ambition of nature policy. The diverse government policy frameworks provided urgency and direction in this regard. As a consequence, policy practice demonstrates an emphasis on the realisation of hectares and potential perverse incentives. No government policy frames are currently available for the two new ambitions, which allows for the opportunity to give substance to these frames in a more shared fashion. There is a need for concrete, flexible goals and relevant, non-rigid indicators. In other words, there is a need for goals and indicators that provide sufficient structure for evaluation for accountability, while simultaneously providing sufficient space for learning along the way and timely adjustments.

- We advise to develop these goals and respective indicators in a shared process with relevant stakeholders (including provincial and national governments and societal parties). Without their involvement, there is a risk that goals and indicators lack ownership and align suboptimal to their practices, increasing the risk for perverse incentives.

We mainly observe accountability towards higher government levels (vertical accountability), and less towards societal partners. Such a more horizontal take on accountability would imply provinces develop policy that takes into account the perspectives and interests of societal parties. As we discussed earlier, limitedly involving societal parties during the evaluation poses a risk. They have not been part of the learning processes and the accompanied shared body of thought that has been

developed regarding the broadened ambitions of nature policy. The staying behind of the learning process of societal parties has consequences for establishing horizontal accountability; their perspectives and interests are limitedly known and may diverge from the current intended policy plans. Additionally, provinces are largely dependent on societal partners for executing nature policy and, thus, the realisation of the ambitions. Horizontal accountability is expected to have a positive effect on goal attainment and, by extension, on vertical accountability.

- As we emphasised earlier, we advise involving societal partners in the following evaluation period and thereby making them part of the development of a shared perspective on the nature policy ambitions. This benefits the production of socially robust knowledge and ownership, but also the balance between learning and accountability. We once more draw attention to the importance of taking the increased complexity as a consequence of the increased number of actors and heterogeneity in perspectives into account in the process design of the evaluation. Next to increased complexity in relation to how the evaluation is organized and dealing with numerous, potentially conflicting, perspectives, attention is also needed for encouraging willingness to participate and establishing a sufficiently safe environment for sharing experiences and opinions.
- To benefit this safe environment it may be useful to alternately plan homo- and heterogeneous workshop or sessions. Homogeneous groups tend to feel more safe, allowing for familiarisation with the evaluation and the researchers, paving the path for following heterogeneous sessions.

The Natuurpact reflexive evaluation has resulted in promising outcomes and has initiated mutual learning between provinces, national government and societal partners in a process of knowledge co-creation. The coming years this process may be continued, further enriching policy practice with socially robust knowledge, supporting the development of a learning policy network and, overall, increasing policy impact on attaining the broadened nature policy ambition by 2027.

References

- Argyris, C., & Schon, D. (1978). *Organizational learning: A theory of action perspective*. Reading, Mass.: Addison Wesley.
- Arkesteijn, M., van Mierlo, B., & Leeuwis, C. (2015). The need for reflexive evaluation approaches in development cooperation. *Evaluation*, 21(1), 99–115.
- Belcher, B. M., Rasmussen, K. E., Kemshaw, M. R., & Zornes, D. A. (2016). Defining and assessing research quality in a transdisciplinary context. *Research Evaluation*, 25(1), 1–17. <https://doi.org/10.1093/reseval/rvv025>
- Bracken, L. J., Bulkeley, H. A., & Whitman, G. (15AD). Transdisciplinary research: understanding the stakeholder perspectives. *Journal of Environmental Planning and Management*, 58(7), 1291–1308.
- Ebrahim, A. (2005). *Accountability Myopia: Losing Sight of Organizational Learning*. *Nonprofit and Voluntary Sector Quarterly* (Vol. 34). <https://doi.org/10.1177/0899764004269430>
- 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>
- Elzen, B., Augustyn, A., Barbier, M., & Mierlo, B. Van. (2017). AgroEcological Transitions: Changes and Breakthroughs in the Making, 336. <https://doi.org/http://dx.doi.org/10.18174/407609>
- Guijt, I. (2010). Accountability and Learning: Exploding the Myth of Incompatibility between Accountability and Learning. In J. Ubels, N.-A. Acquaye-Baddoo, & A. Folwer (Eds.), *Capacity Development in Practice* (pp. 277–291). Earthscan. Retrieved from <http://www.snvworld.org/en/publications/capacity-development-in-practice-complete-publication>
- Jahn, T., & Keil, F. (2015). An actor-specific guideline for quality assurance in transdisciplinary research. *Futures*, 65, 195–208. <https://doi.org/10.1016/j.futures.2014.10.015>
- Kirkhart, K. E. (2000). Reconceptualizing Evaluation Use: An Integrated Theory of Influence. *New Directions for Evaluation*, (88), 5.
- Klerkx, L., Seuneke, P., de Wolf, P., & Rossing, W. A. H. (2017). Replication and translation of co-innovation: The influence of institutional context in large international participatory research projects. *Land Use Policy*, 61, 276–292. <https://doi.org/10.1016/j.landusepol.2016.11.027>
- Lang, D. J., Wiek, A., Bergmann, M., Stauffacher, M., Martens, P., Moll, P., ... Thomas, C. J. (2012). Transdisciplinary research in sustainability science: Practice, principles, and challenges. *Sustainability Science*, 7(SUPPL. 1), 25–43. <https://doi.org/10.1007/s11625-011-0149-x>
- Mattor, K., Betsill, M., Huayhuaca, C., Huber-Stearns, H., Jedd, T., Sternlieb, F., ... Luizza, M. (2013). Transdisciplinary Research on Environmental Governance: A View from the Trenches. *Earth System Governance Working Papers*, 29(October), 26. <https://doi.org/10.1016/j.envsci.2014.06.002>
- Mausser, W., Klepper, G., Rice, M., Schmalzbauer, B. S., Hackmann, H., Leemans, R., & Moore, H.

- (2013). Transdisciplinary global change research: the co-creation of knowledge for sustainability. *Current Opinion in Environmental Sustainability*, 5(3–4), 420–431. <https://doi.org/10.1016/j.cosust.2013.07.001>
- Meer, van der F.-B., & Edelenbos, J. (2006). Evaluation in Multi-Actor Policy Processes: Accountability, Learning and Co-operation. *Evaluation*, 12(2), 201–218. <https://doi.org/10.1177/1356389006066972>
- Nowotny, H. (2000). Re-thinking Science : From Reliable Knowledge to Societally Robust Knowledge. In *Jahrbuch 2000 des Collegium Helveticum* (pp. 1–19). Zurich.
- Patton, M. Q. (2000). Utilization-Focused Evaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation Models* (pp. 425–438). Boston: Kluwer Academic Publishers.
- Patton, M. Q. (2008). *Utilization-Focused Evaluation*. Sage Publications, Inc.
- Preskill, H., & Torres, R. (2000). The Learning Dimension of Evaluation Use. *New Directions for Evaluation*, 2000(88), 25–37. <https://doi.org/10.1002/ev.1189>
- Regeer, B., De Wildt-Liesveld, R., van Mierlo, B., & Bunders, J. F. G. (2016). Exploring ways to reconcile accountability and learning in the evaluation of niche experiments. *Evaluation*, 22(1), 6–28.
- Regeer, B. J., Hoes, a.-C., van Amstel-van Saane, M., Caron-Flinterman, F. F., & Bunders, J. F. G. (2009). Six Guiding Principles for Evaluating Mode-2 Strategies for Sustainable Development. *American Journal of Evaluation*, 30(4), 515–537. <https://doi.org/10.1177/1098214009344618>
- Richmond, B. J., Mook, L., & Quarter, J. (2003). Societal accounting for nonprofits: Two models. *Nonprofit Management and Leadership*, 13(4), 308–324.
- Sarkki, S., Heikkinen, H. I., & Karjalainen, T. P. (2013). Sensitivity in transdisciplinary projects: A case of reindeer management in Finland. *Land Use Policy*, 34, 183–192. <https://doi.org/10.1016/j.landusepol.2013.03.004>
- Schön, D. A., & Rein, M. (1994). *Frame Reflection*. New York: Basic Books Inc.
- Shulha, L. M., Whitmore, E., Cousins, J. B., Gilbert, N., & al Hudib, H. (2015). Introducing Evidence-Based Principles to Guide Collaborative Approaches to Evaluation: Results of an Empirical Process. *American Journal of Evaluation*, 37(2), 193–215. <https://doi.org/10.1177/1098214015615230>
- 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*.
- Veen, van S. C., de Wildt-liesveld, R., Bunders, J. F. G., & Regeer, B. J. (2014). Supporting reflective practices in societal change processes with the dynamic learning agenda : an example of learning about the process towards disability inclusive development Saskia C. van Veen *,. *Int. J. Learning and Change*, 7(3/4), 211–233.
- Walter, A. I., Helgenberger, S., Wiek, A., & Scholz, R. W. (2007). Measuring societal effects of transdisciplinary research projects: design and application of an evaluation method. *Evaluation and Program Planning*, 30(4), 325–38. <https://doi.org/10.1016/j.evalprogplan.2007.08.002>

