



Rapport

Evaluatieonderzoek subsidieregeling Nationaal Groeifonds

Auteurs

Ir. Arthur Vankan

Dr. Ir. Matthijs Janssen

Ir. Emma Urselmann

Ir. Menno Driesse

Rapport

Evaluatieonderzoek subsidieregeling Nationaal Groeifonds

Auteurs

Ir. Arthur Vankan
Dr. Ir. Matthijs Janssen
Ir. Emma Urselmann
Ir. Menno Driesse

Opdrachtgever

Ministerie van Economische Zaken

Publicatienummer

2024.118-2449

Datum

26 november 2024

Beeld omslag

Freepik

Inhoud

Managementsamenvatting	4
1 Introductie	11
1.1 Aanleiding	11
1.2 Onderzoeksvragen	11
1.3 Leeswijzer	12
2 De NGF-subsidieprojecten	14
2.1 Inleiding	14
2.2 Groeien met Groen Staal	15
2.3 Charging Energy Hubs	19
3 Evaluatiebegrippen en -mogelijkheden	24
3.1 Bouwstenen voor evaluatiemethoden	24
3.2 Verdieping van kansrijke evaluatieaanpakken	30
4 Methode voor evaluatie NGF-projecten	42
4.1 Logische selectie van passende evaluatieaanpakken	42
4.2 Stappen voor het toepassen van de voorgestelde methode	44
4.3 Geschiktheid van voorgestelde methode	49
5 Evaluatieontwerp Groeien met Groen Staal	51
5.1 Inleiding	51
5.2 Doeltreffendheid GGS	51
5.3 Doelmatigheid GGS	68
5.4 Vooruitblik	72
6 Evaluatieontwerp Charging Energy Hubs	76
6.1 Inleiding	76
6.2 Doeltreffendheid CEH	76
6.3 Doelmatigheid CEH	95
6.4 Vooruitblik	99
7 Aandachtspunten & aanbevelingen	103
7.1 Evaluatiemethode	103
7.2 Aandachtspunten & aanbevelingen	105
7.3 Overzicht aanbevelingen	110
7.4 Slotwoord	111

Managementsamenvatting

Aanleiding (H1)

Voor de **Subsidieregeling Nationaal Groeifonds (NGF)** is een evaluatieplan verplicht vanuit de Europese Commissie (EC) om gebruik te kunnen maken van een aantal steuncategorieën van de Algemene groepsvrijstellingsverordening (AGVV). Op 28 augustus 2023 is het door de EC goedgekeurde evaluatieplan met de Tweede Kamer gedeeld. Met de Europese Commissie is afgesproken dat in 2024 al een eerste rapportage zal worden aangeleverd over de met de Subsidieregeling Nationaal Groeifonds gefinancierde NGF-projecten. In 2024 zijn die gefinancierde projecten onvoldoende gevorderd om een evaluatie te doen naar de doeltreffendheid en de doelmatigheid daarvan. Wel kan inzicht worden gegeven in de kenmerken van de subsidieverleningen (welke activiteiten, hoogte subsidiebedrag, eindbegunstigden, verplichtingen etc.). Ook kan - op basis van de gefinancierde projecten - een nadere uitspraak worden gedaan over de kansrijkheid van evaluatietechnieken voor vervolgevaluaties en inzicht worden gegeven in de beschikbare data. Daarna zal de eerste vijfjaarlijkse evaluatie van het Nationaal Groeifonds in 2026 aan de Europese Commissie worden voorgelegd.

Het door de EC goedgekeurde evaluatieplan is opgesteld t.b.v. het hele Nationaal Groeifonds, dus ziet naast de subsidieroute ook toe op de departementale route. Om die reden **worden er in dit onderzoek ook aanbevelingen gedaan over de wijze waarop de departementale projecten kunnen worden geëvalueerd**. De algehele evaluatie van het Nationaal Groeifonds zal worden vormgegeven als een meta-evaluatie van de evaluaties van de 50 individuele NGF-projecten.

De twee NGF-subsidieprojecten (H2)

Het project **Groeien met Groen Staal** richt zich op het verduurzamen van de Nederlandse staalindustrie door tegen 2050 een CO₂-neutraal productieproces te realiseren. Het project kent een totaalbudget van 126 miljoen euro, waarvan 101 miljoen euro gefinancierd wordt vanuit het Nationaal Groeifonds. Via vijf werkpakketten en demonstrators worden oplossingen getest om emissies te verminderen en de overgang naar circulaire, groene staalproductie technisch en economisch haalbaar te maken.

Het project **Charging Energy Hubs** heeft als doel de uitrol van oplaadinfrastructuur voor elektrische logistieke voertuigen te versnellen en zo de belasting op het elektriciteitsnetwerk te verlichten. Het project kent een totaalbudget van ruim 70 miljoen euro, waarvan 39 miljoen euro gefinancierd wordt vanuit het Nationaal Groeifonds. Door middel van 10 werkpakketten, waaronder de ontwikkeling van slimme energiesystemen en praktijktests op verschillende locaties, optimaliseert het project de integratie van oplaadinfrastructuur en hernieuwbare energiebronnen in Nederland.

In onderstaande tabel wordt samenvattende informatie gegeven over (a) de verleende subsidiebedragen, (b) de eindbegunstigden voor de subsidie, (c) de activiteiten

waarvoor subsidie is verleend en (d) de verplichtingen die verbonden zijn aan de subsidiebeschikking.

Deelvraag	Groeien met Groen Staal	Charging Energy Hubs
a. Welke subsidiebedragen zijn toegekend?	€101 miljoen	€38,9 miljoen
b. Wie zijn de eindbegunstigden voor deze subsidie?	32 partijen: Zie Tabel 2	33 partijen: Zie Tabel 3
c. Voor welke activiteiten zijn de subsidies verleend?	Experimentele ontwikkeling: 36% Fundamenteel onderzoek: 28% Bouw onderzoeksinfrastructuur: 17% Industrieel onderzoek: 12% Niet economische activiteiten: 4% Haalbaarheidsstudies: 3%	Experimentele ontwikkeling: 56% Industrieel onderzoek: 29% Niet economische activiteiten: 14% Opleidingssteun: 0,3%
d. Welke verplichtingen zijn verbonden aan de subsidiebeschikking?	1) Algemene NGF-voorwaarden ¹ 2) Project-specifieke verplichtingen omtrent: - Looptijd - Opsturen samenwerkingsovereenkomst - Opsturen rapportage integrale kostensystematiek (NHL Stenden) - Maatwerkafspraken Tata Steel	1) Algemene NGF-voorwaarden ² 2) Project-specifieke verplichtingen omtrent: - Opsturen samenwerkingsovereenkomst - Opsturen rapportage integrale kostensystematiek (Damen Research)

Een relatief universele methode voor het evalueren van NGF-projecten (H3+H4)

Voor uiteenlopende beleidsevaluaties wordt een breed palet aan evaluatiemethoden ingezet. In hoofdstuk 3 in dit rapport gaan we hier in detail op in en beschrijven we verschillende begrippen zoals evaluatiedoelen, evaluatiestromingen-/tradities, evaluatietypen, evaluatieaanpakken, analysetechnieken, en dataverzamelingstechnieken.

We concluderen dat voor het evalueren van de doeltreffendheid en doelmatigheid van deze twee NGF-projecten een **combinatie van 'contribution analysis' en 'process tracing' (CA/PT)** een robuuste basis biedt. De gecombineerde aanpak is krachtig in een context waarin [1] er sprake is van parallelle actielijnen en veel (verschillende typen)

¹ RVO (2024) *Handleiding subsidieregeling NGF*. Geraadpleegd op 4 oktober 2024. <https://www.rvo.nl/sites/default/files/2023-11/Handleiding%20subsieregeling%20Nationaal%20Groeifonds%202024-2025.pdf>

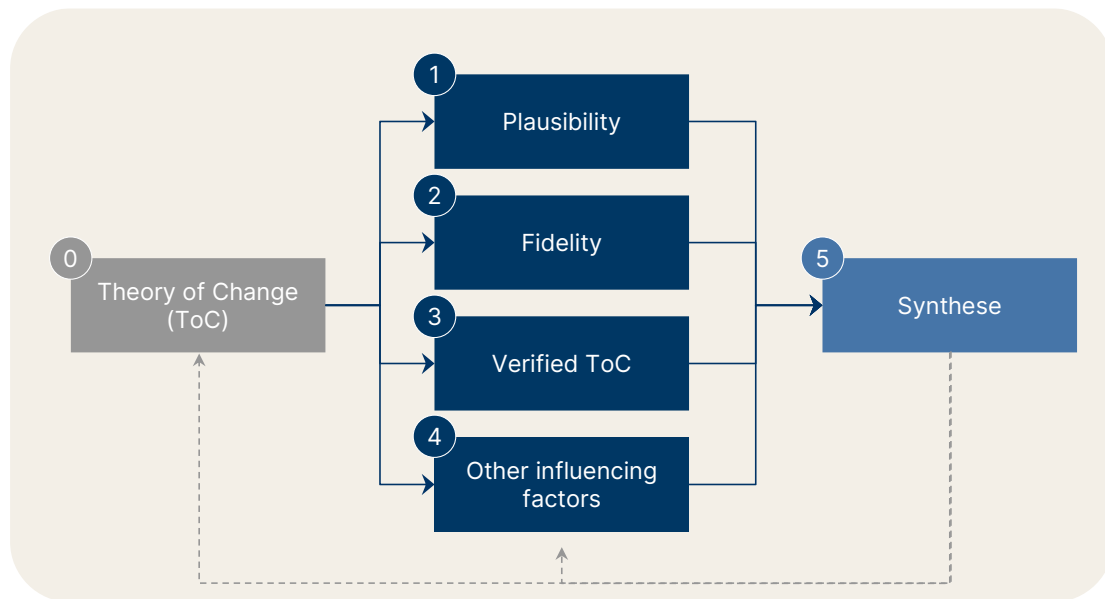
² RVO (2024) *Handleiding subsidieregeling NGF*. Geraadpleegd op 4 oktober 2024. <https://www.rvo.nl/sites/default/files/2023-11/Handleiding%20subsieregeling%20Nationaal%20Groeifonds%202024-2025.pdf>

betrokken partijen, [2] er niet gewerkt kan worden met duidelijke experimentele groepen en controlegroepen op het niveau van het gehele NGF-project, [3] beoogde uitkomsten niet altijd even goed te kwantificeren zijn, en [4] er een Theory of Change³ ligt die als basis kan dienen voor de evaluatie. Deze aanpak lijkt om deze redenen ook geschikt voor (de meeste) andere NGF-projecten. Een aanpak die volledig leunt op het bestaan van een counterfactual (op projectniveau) lijkt niet haalbaar in deze context. Alternatieve aanpakken zoals outcome mapping/harvesting kunnen weliswaar toegepast worden, maar zijn minder toegespitst op het onderzoeken van (beoogde) causale relaties zoals die in een Theory of Change zijn opgesteld. Alle opties meewegende kan de CA+PT-aanpak zowel voor de projecten via de subsidieroute als voor de projecten via de departementale route als basis genomen worden. Voor de projecten in de departementale route moet wel verder verkend worden of de Theories of Change voldoende uitgewerkt zijn en er voldoende data en informatie beschikbaar zijn om een dergelijke aanpak goed uit te kunnen voeren.

Binnen een contribution analysis dient de opgestelde Theory of Change in het NGF-projectplan als basis. Vervolgens worden achtereenvolgens vier criteria getoetst om de beoogde causale relaties zoals beschreven in de Theory of Change te beoordelen⁴: [1] de '**plausibility**' van de Theory of Change c.q. is de Theory of Change goed onderbouwd en uitgewerkt en daarmee geloofwaardig, [2] de '**Fidelity**' c.q. worden de activiteiten ook daadwerkelijk uitgevoerd zoals beschreven in de Theory of Change, [3] een '**Verified** Theory of Change' c.q. het verzamelen van bewijs voor het wel/niet bestaan van de causale relaties, en [4] '**other influencing factors**' c.q. andere factoren die invloed kunnen hebben op de uitkomsten zoals beschreven in de Theory of Change. Dit vierde criterium zorgt er onder meer voor dat een eventuele 'positieve bias' (of zelfs wensdenken) vanuit subsidieaanvragers niet onbewust in de evaluatie overgenomen wordt. Tot slot worden alle bevindingen bijeengebracht om een onderbouwde 'contribution story' op te stellen waarin aandacht is voor de verschillende impactpaden in de Theory of Change.

³ De beste Nederlandse vertaling voor een Theory of Change is een 'beleidstheorie'. In dit rapport hanteren wij de Engelse term, omdat deze de afgelopen jaren binnen de context van het NGF gebruikt is en omdat dit aansluit bij internationale literatuur.

⁴ Ook voor de vier criteria zullen we in dit rapport de Engelse termen hanteren.



Process tracing is een krachtige manier om het derde criterium binnen de Contribution Analysis ('Verified Theory of Change') concreet te toetsen. Met process tracing kan getoetst worden of de causale relaties zoals geschetst in de Theory of Change ook daadwerkelijk bestaan. Process tracing doet dit door de veronderstelde relaties te verifiëren of te falsificeren. Process tracing wordt concreet invulling gegeven door het ontwikkelen en toepassen van vier typen 'process tracing testen'. Deze vier typen testen verschillen enerzijds in de mate waarin ze bewijskracht leveren voor de onderzochte causale relatie en anderzijds in de mate waarin ze uitvoerbaar zijn in de praktijk. In generieke zin geldt dat een sterke test met veel bewijskracht voor het onderbouwen van een causale relatie doorgaans moeilijker te ontwikkelen en uit te voeren is.

De evaluatiemethode toegepast op de twee NGF-subsidieprojecten (H5+H6)

De beschreven evaluatiemethode is concreet uitgewerkt voor de twee NGF-projecten Groeien met Groen Staal (H5) en Charging Energy Hubs (H6). Het resultaat is een **concreet evaluatieontwerp** dat ingezet kan worden om de doeltreffendheid en doelmatigheid van deze projecten te evalueren. Voor ieder project zijn achtereenvolgens de bovenstaande criteria uitgewerkt. Daarbij is ook **in detail in kaart gebracht welke data beschikbaar zijn** per onderdeel ('blokje') in de Theory of Change, en zijn vervolgens process tracing tests ontwikkeld voor alle impactpaden ('pijltjes') in de Theory of Change. De doelmatigheid kan onderzocht worden door eerst te kijken naar [1] de uitvoeringsdoelmatigheid (subsidie, uitvoering en uitvoeringskosten, en administratieve lasten), dan [2] de microdoelmatigheid (relatie opbrengsten op projectniveau c.q. outputs en gemaakte kosten) en tot slot [3] de macrodoelmatigheid (relatie outcomes en impact en gemaakte kosten). De evaluatieontwerpen sluiten af met een vooruitblik ten aanzien van wat mogelijk is in de evaluaties van het Nationaal Groeifonds in 2026 en daarna. Een belangrijk element is dat er beperkingen worden opgeworpen door het moment in tijd waarop geëvalueerd wordt. In 2026 zijn de projecten nog in volle gang en hebben nog niet alle outcomes en impacts de kans gehad om zich te manifesteren.

De evaluaties in 2026 zullen het karakter hebben van tussentijdse evaluaties waarbij de focus ligt op de inbreng van subsidieontvangers ('input'), voortgang van de voorgenomen activiteiten ('throughput'), en de eerste projectresultaten ('outputs'). Mogelijk zullen er eerste tekenen zijn van eerste-orde-effecten ('intermediate outcomes'). Naarmate de projecten verder gevorderd zijn kan ook meer gemeten en geëvalueerd worden. Een kanttekening bij Charging Energy Hubs is daarbij dat de verwachte tweede-orde-effecten ('outcomes') en uiteindelijke impact zich pas ná afloop van de projectdoorlooptijd zullen manifesteren en dat het daarmee theoretisch onmogelijk is om deze outcomes en impacts op een empirische manier te evalueren direct na afronding van het project. Voor Groeien met Groen Staal speelt dit in mindere mate, omdat vrijwel alle benodigde indicatoren beschikbaar moeten komen in het jaar waarin het project wordt afgerond (2031).

Indien men toch een evaluatie wil uitvoeren zonder dat de volledige Theory of Change zich heeft kunnen manifesteren, kan men de empirie met de theorie combineren. In een dergelijke situatie kunnen de indicatoren en de process tracing tests tot zover mogelijk uitgevoerd worden ten behoeve van een **empirische onderbouwing** (bijv. tot en met de intermediate outcomes). Voor het resterende 'rechterdeel' in de Theory of Change (outcomes en impact) kan een **theoretische onderbouwing** conform de Theory of Change geboden worden. Voor dit laatste geldt dat de plausibiliteit van (de resterende onderdelen van) de Theory of Change essentieel is, omdat men op basis van de theorie moet 'aannemen' dat de effecten in de toekomst zullen optreden.

Aandachtspunten en aanbevelingen (H7)

In het slothoofdstuk wordt een aantal algemene aandachtspunten en aanbevelingen geboden. De **belangrijkste aanbevelingen zijn op de volgende pagina weergegeven**.

De voorgestelde evaluatiemethode voor NGF-projecten biedt een solide basis voor toekomstige evaluaties en kan in feite voor ieder NGF-project ingezet worden. **De uitdaging zal met name zitten in het geven van invulling aan deze aanpak**. Voor de departementale projecten zal het geven van invulling aan de CA/PT-aanpak mogelijk uitdagender zijn, gelet op de complexiteit en omvang van deze projecten. In dit rapport worden hiervoor reeds enkele aanbevelingen gedaan gelet op toekomstige evaluaties.

Aanbevelingen

Specifieke aanbevelingen voor Groeien met Groen Staal + Charging Energy Hubs

1. Ga als RVO en consortia in gesprek over de benoemde indicatoren in het projectplan, en maak duidelijke afspraken (bij voorkeur per indicator) over of en hoe deze verzameld gaan worden en hoe hierover gerapporteerd gaat worden in de voortgangsrapportages;
2. Voer als EZ (en RVO) voorafgaand aan de evaluaties een data assessment uit, en houd er als toekomstige evaluatoren rekening mee dat er (op delen van) het hier gepresenteerde voorstel voor een evaluatieontwerp een plan B voor dataverzameling ontwikkeld moet worden;
3. Probeer als toekomstige evaluatoren (en eventueel als EZ en RVO) de gepresenteerde process tracing testen te verbeteren met als doel het verder verstevigen van de robuustheid van het evaluatieontwerp.

Algemene aanbevelingen voor evaluaties van NGF-projecten

4. Scherp als verantwoordelijke departementen en NGF-consortia vroegtijdig de Theory of Change aan gezien de relevantie hiervan voor een robuuste evaluatie;
5. Verken als verantwoordelijke departementen of er op basis van generieke databronnen specifieke indicatoren voor meerdere NGF-projecten ontwikkeld kunnen worden;
6. Streef als toekomstige evaluatoren bij het trekken van conclusies over de doeltreffendheid in de complexe contexten van NGF-projecten naar een goede balans tussen eenvoud en uitlegbaarheid enerzijds en nuance anderzijds;
7. Houd er bij het uitvoeren van de NGF-evaluaties op projectniveau en de algehele evaluatie van het NGF rekening mee dat het evalueren van de gehonoreerde NGF-projecten een niet-volledig beeld geeft van de mogelijke effecten van de NGF-subsidie en het NGF als geheel;

Hoofdstuk 1

Introductie

In dit hoofdstuk wordt de aanleiding van dit onderzoek beschreven en worden de onderzoeksvragen gepresenteerd.

1 Introductie

1.1 Aanleiding

Voor de **Subsidieregeling Nationaal Groeifonds (NGF)** is een **evaluatieplan verplicht vanuit de Europese Commissie (EC)** om gebruik te kunnen maken van een aantal steuncategorieën van de Algemene groepsvrijstellingsverordening (AGVV). Op 28 augustus 2023 is het door de EC goedgekeurde evaluatieplan met de Tweede Kamer gedeeld. Met de Europese Commissie is afgesproken dat in 2024 al een eerste rapportage zal worden aangeleverd over de met de Subsidieregeling Nationaal Groeifonds gefinancierde NGF-projecten. In 2024 zijn die gefinancierde projecten onvoldoende gevorderd om een evaluatie te doen naar de doeltreffendheid en de doelmatigheid daarvan. Wel kan inzicht worden gegeven in de kenmerken van de subsidieverleningen (welke activiteiten, hoogte subsidiebedrag, eindbegunstigden, verplichtingen etc.). Ook kan - op basis van de gefinancierde projecten - een nadere uitspraak worden gedaan over de kansrijkheid van evaluatietechnieken voor vervolgevaluaties en inzicht worden gegeven in de beschikbare data. Daarna zal de eerste vijfjaarlijkse evaluatie van het Nationaal Groeifonds in 2026 aan de Europese Commissie worden voorgelegd.

Het door de EC goedgekeurde evaluatieplan is opgesteld t.b.v. het hele Nationaal Groeifonds, dus ziet naast de subsidieroute ook toe op de departementale route. Om die reden **worden er in dit onderzoek ook aanbevelingen gedaan over de wijze waarop de departementale projecten kunnen worden geëvalueerd**. De algehele evaluatie van het Nationaal Groeifonds zal worden vormgegeven als een meta-evaluatie van de evaluaties van de 50 individuele NGF-projecten.

1.2 Onderzoeksvragen

Het ministerie van Economische Zaken (EZ) heeft de volgende onderzoeksvragen opgesteld⁵:

1. Het onderzoek moet **inzicht geven in de subsidies** die zijn verleend op de volgende punten:
 - Welke subsidiebedragen zijn toegekend?
 - Wie zijn de eindbegunstigden voor deze subsidies?
 - Voor welke activiteiten zijn deze subsidies verleend?
 - Welke verplichtingen zijn verbonden aan de subsidiebeschikkingen?

⁵ Bron: Programma van Eisen binnen aanbesteding van de opdracht voor het opstellen van dit evaluatierapport

2. Wat zijn voor deze subsidieprojecten mogelijke **kansrijke technieken** voor toekomstige evaluaties van doelmatigheid en doeltreffendheid en kunnen deze vanaf 2026 worden toegepast?
 - Hierbij zullen in ieder geval de technieken worden betrokken die in het evaluatieplan zijn opgenomen, te weten: contribution analysis, process tracing en mogelijk ook benchmarking. De opdrachtgever wordt ook geadviseerd over andere kansrijke technieken voor de evaluatie van systeem- en transitiebeleid⁶.
3. **Welke data zijn er beschikbaar bij RVO** over de projecten die subsidie hebben ontvangen en in hoeverre is dit voldoende om gebruik te maken van de genoemde evaluatietechnieken?
 - Hierbij zal moeten worden betrokken in hoeverre de data die is uitgevraagd op basis van het evaluatieplan beschikbaar is en of de data van voldoende kwaliteit is.
 - Hierbij zal betrokken moeten worden in hoeverre de data die wordt uitgevraagd voldoende basis biedt om gebruik te maken van de evaluatietechnieken.

1.3 Leeswijzer

Dit evaluatierapport is als volgt opgebouwd:

- In **hoofdstuk 2** worden de twee subsidieprojecten Groeien met Groen Staal en Charging Energy Hubs beschreven, en wordt inzicht gegeven in de subsidieverlening (onderzoeksvraag 1).
- In **hoofdstuk 3** worden enkele belangrijke evaluatiebegrippen en de relatie daartussen beschreven, en wordt nader ingegaan op diverse soorten theorie-gebaseerde evaluatie-aanpakken die we relevant achten in de NGF-context van zowel de subsidieroute als de departementale route.
- In **hoofdstuk 4** wordt een voorstel gedaan voor een relatief universele methode voor het evalueren van NGF-projecten voor projecten in zowel de subsidieroute als de departementale route.
- In **hoofdstuk 5** wordt een evaluatieontwerp voor NGF-subsidieproject 'Groeien met Groen Staal' gepresenteerd.
- In **hoofdstuk 6** wordt een evaluatieontwerp voor NGF-subsidieproject 'Charging Energy Hubs' gepresenteerd.
- In **hoofdstuk 7** worden overkoepelende aandachtspunten en aanbevelingen gepresenteerd voor toekomstige evaluaties van zowel de twee subsidieprojecten als NGF-projecten in de departementale route.

⁶ Zie hiervoor het rapport *Durf te leren, ga door met meten* van de Commissie Evaluatiemethoden Systeem- en Transitiebeleid.

Hoofdstuk 2

De NGF-subsidieprojecten

In dit hoofdstuk worden de twee NGF-subsidieprojecten 'Groeien met Groen Staal' en 'Charging Energy Hubs' beschreven.



2 De NGF-subsidieprojecten

2.1 Inleiding

De subsidieregeling NGF heeft in 2023 subsidie verleend aan twee projecten, Groeien met Groen Staal en Charging Energy Hubs. In de volgende paragrafen wordt per project antwoord gegeven op de volgende 4 deelvragen:

- Welke subsidiebedragen zijn toegekend?
- Wie zijn de eindbegunstigden voor deze subsidie?
- Voor welke activiteiten is de subsidie verleend?
- Welke verplichtingen zijn verbonden aan de subsidiebeschikking?

Onderstaande tabel geeft een samenvatting van de antwoorden op deze deelvragen:

*Tabel 1: Overzicht beantwoording deelvragen. * O&O&I = Onderzoek, Ontwikkeling en Innovatie*

Deelvraag	Groeien met Groen Staal	Charging Energy Hubs
a. Welke subsidiebedragen zijn toegekend?	€101 miljoen	€38,9 miljoen
b. Wie zijn de eindbegunstigden voor deze subsidie?	32 partijen: Zie Tabel 2	33 partijen: Zie Tabel 3
c. Voor welke activiteiten zijn de subsidies verleend?	Experimentele ontwikkeling: 36% Fundamenteel onderzoek: 28% Bouw onderzoeksinfrastructuur: 17% Industrieel onderzoek: 12% Niet economische activiteiten: 4% Haalbaarheidsstudies: 3%	Experimentele ontwikkeling: 56% Industrieel onderzoek: 29% Niet economische activiteiten: 14% Opleidingssteun: 0,3%
d. Welke verplichtingen zijn verbonden aan de subsidiebeschikking?	1) Algemene NGF-voorwaarden 2) Project-specifieke verplichtingen omtrent: - Looptijd - Opsturen samenwerkingsovereenkomst - Opsturen rapportage integrale kostensystematiek (NHL Stenden) - Maatwerkafspraken Tata Steel	1) Algemene NGF-voorwaarden 2) Project-specifieke verplichtingen omtrent: - Opsturen samenwerkingsovereenkomst - Opsturen rapportage integrale kostensystematiek (Damen Research)

2.2 Groeien met Groen Staal



2.2.1 Omschrijving project

Het achtjarige programma **Groeien met Groen Staal** heeft vanuit het NGF meer dan 100 miljoen euro aan financiering ontvangen. Het project heeft als doel om de Nederlandse staalindustrie te verduurzamen. Daarmee richt het zich in het kader van het Klimaatakkoord vooral op het behalen van het klimaatbeleid binnen de staalsector. Daarnaast sluit het project aan bij de beleidsdoelen rondom het verbeteren van het leefmilieu en het streven naar (grondstoffen)autonomie.

De belangrijkste doelstelling van het project is om in 2050 een CO₂-neutraal staalproductieproces gerealiseerd te hebben. Hierbij richt het project zich op het gehele productieproces van groen staal om de emissies van CO₂, NO_x en fijnstof significant te verminderen: van de aanmaak van groene staalproducten tot het recyclen van de staalschroot. Om deze doelen te behalen bestaat het project uit een vijftal werkpakketten. Deze werkpakketten bevatten zogenoemde 'demonstrators' die oplossingen stimuleren voor de transformatie naar een circulaire groene staalsector, waarbij niet alleen de technische haalbaarheid, maar ook de economische haalbaarheid op productieniveau worden onderzocht.

- 1) *Systeemverandering*: Organisatorische veranderingen en systeemveranderingen die nodig zijn om de staalindustrie te verduurzamen (o.a. beoordelen van de impact van het overheidsbeleid en externe factoren).
 - 1 demonstrator in onderdeel van dit werkpakket.
- 2) *Productie*: Ontwikkeling van technologieën voor de decarbonisatie van staalproductie (o.a. op basis van waterstof en hernieuwbare energie).
 - 4 demonstrators zijn onderdeel van dit werkpakket.
- 3) *Verwerking*: Onderzoek naar verwerkingsprocessen voor staal met een andere samenstelling (door nieuwe productiemethode).
 - 5 demonstrators zijn onderdeel van dit werkpakket.
- 4) *Gebruik*: Onderzoek naar de prestaties van groen staal (o.a. hardheid, sterkte, bestendigheid).
 - 4 demonstrators zijn onderdeel van dit werkpakket.
- 5) *Terugwinning*: Optimaliseren van terugwinning van staal uit schroot en het voorkomen van kwaliteitsverlies bij recycling.
 - 4 demonstrators zijn onderdeel van dit werkpakket.

Het project bestaat uit een consortium van ruim 30 partners. Hierin werken onder andere verschillende onderwijs- en kennisinstellingen samen met industriële partners (zoals Tata Steel).

2.2.2 Toegekende subsidiebedragen en eindbegunstigden

Groeien met Groen Staal heeft in totaal een budget van €126 miljoen, waarvan €101 miljoen gefinancierd wordt vanuit het NGF. Tabel 2 geeft aan hoe de subsidie is verdeeld over de deelnemende partijen:

Tabel 2: Subsidiebedrag GGS per deelnemer. Bron: Verleningsbrief Groeien met Groen Staal

Deelnemer	Bedrag	% van subsidie
Technische Universiteit Delft	€ 22.769.171	22,5%
Tata Steel Nederland Technology B.V.	€ 15.043.666	14,9%
Technische Universiteit Eindhoven	€ 12.862.920	12,7%
Universiteit Twente	€ 8.005.216	7,9%
Rijksuniversiteit Groningen	€ 7.873.956	7,8%
Stichting Materials innovation institute (M2i)	€ 6.077.000	6,0%
Ex Mente Technologies (Pty) Ltd	€ 3.445.940	3,4%
SKF B.V.	€ 3.216.522	3,2%
Philips Consumer Lifestyle B.V.	€ 2.577.218	2,6%
Nederlandse Organisatie voor toegepast-natuurwetenschappelijk onderzoek TNO	€ 2.000.071	2,0%
Bosch Transmission Technology B.V.	€ 1.872.000	1,9%
Rijksuniversiteit Leiden	€ 1.567.980	1,6%
Universiteit Utrecht	€ 1.287.936	1,3%
Space XYZ	€ 1.156.125	1,1%
Alleima Benelux B.V.	€ 1.145.430	1,1%
Stichting Avans	€ 1.055.553	1,0%
NHL Stenden	€ 871.328	0,9%
ARN Auto Recycling Nederland B.V.	€ 854.875	0,8%
Stichting voor educatie en beroepsonderwijs Nova College	€ 800.000	0,8%
DevControl	€ 768.000	0,8%
Researchable	€ 768.000	0,8%
In Summa Innovation B.V.	€ 720.923	0,7%
VDL Weweler	€ 690.000	0,7%
Nederlandse Onttinningsfabriek	€ 632.037	0,6%
KE Works B.V.	€ 576.000	0,6%

Deelnemer	Bedrag	% van subsidie
Max Planck Institut für Eisenforschung	€ 554.400	0,5%
FNsteel B.V.	€ 494.813	0,5%
Bouwen met Staal	€ 396.540	0,4%
Koninklijke Nedschroef Holding B.V.	€ 317.824	0,3%
AG der Dillinger Hüttenwerke	€ 309.063	0,3%
Volvo Cars Corporation	€ 166.795	0,2%
SGS Intron	€ 125.681	0,1%
CMT International B.V. ⁷	-	-
Totaal	€ 101.002.983	100%

2.2.3 Voor welke activiteiten zijn de subsidie verleend?

Zoals te lezen in 2.2.1 bestaat het programma uit een groot aantal werkpakketten en demonstrators. Deze activiteiten worden met mede behulp van het groeifonds gefinancierd.

Volgens de voorwaarden van het NGF moeten activiteiten vallen onder 1) kennisontwikkeling, en/of 2) Onderzoek, Ontwikkeling en Innovatie (O&O&I)⁸. Voor Groeien met Groen Staal geldt volgens de ingediende projectbegroting dat de activiteiten voor 33% vallen onder kennisontwikkeling en voor 67% onder O&O&I. Deze verdeling kan verklaard worden door de aard van het project, dat naast technologische innovaties ook een sterke focus heeft op de strategische transformatie van de staalindustrie.

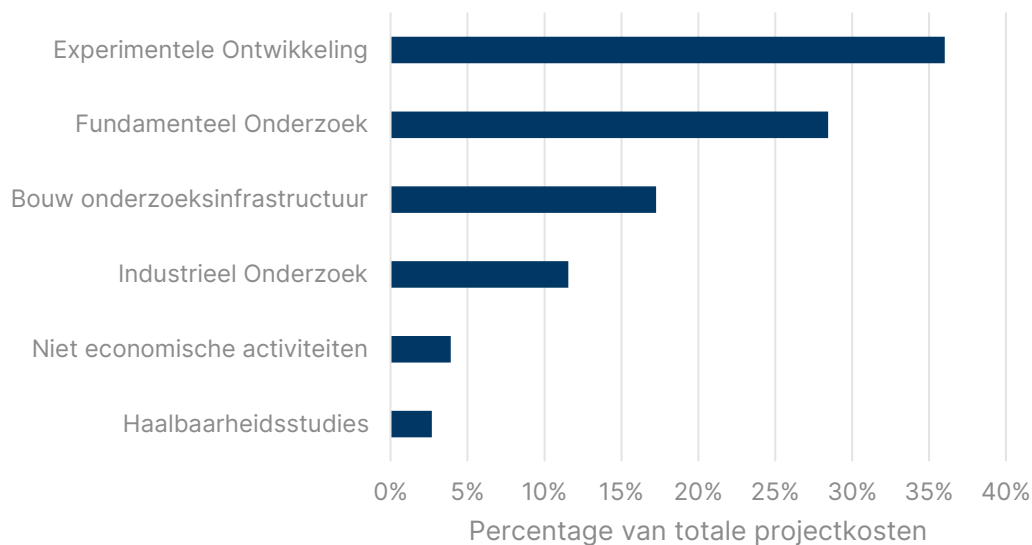
Er geldt een afbakening voor activiteiten die binnen de subsidiabele activiteiten van het NGF vallen. Artikelen 25 t/m 29 en 31 van de AGVV zijn hierbij leidend. De totale projectkosten zijn volgens de ingediende projectbegroting als volgt verdeeld over de desbetreffende categorieën^{9 10}:

⁷ CMT International B.V. voldeed bij de beoordeling niet aan voorwaarde van de subsidieregeling, het subsidiebedrag is daarom op nihil gesteld.

⁸ RVO (2024) *Handleiding subsidieregeling NGF*. Geraadpleegd op 4 oktober 2024. <https://www.rvo.nl/sites/default/files/2023-11/Handleiding%20subsidieregeling%20Nationaal%20Groeifonds%202024-2025.pdf>

⁹ De activiteiten die vallen onder exploitatie innovatieclusters, innovatie klein en middelgrote ondernemingen, proces- en organisatie innovatie en opleidingssteun zijn niet opgenomen in de figuur omdat de kosten hiervan nihil zijn.

¹⁰ Deze uitsplitsing is gebaseerd op de projectbegroting die voor de verlening is opgesteld. Bij de verlening zijn er een aantal wijzigingen doorgevoerd en is de hoogte van de subsidie aangepast wat ertoe kan leiden dat de percentages per activiteit na verlening veranderd zijn.



Figuur 1: Totale projectkosten per categorie activiteiten GGS

2.2.4 Welke verplichtingen zijn verbonden aan de subsidiebeschikking?

Voor Groeien met Groen Staal gelden de algemene NGF-voorwaarden waaraan moet worden voldaan, zie hiervoor de website van de RVO¹¹. Alle deelnemende partijen zijn aan de voorgeschreven verplichtingen gebonden. Indien een partij zich niet aan de verplichtingen houdt, kan dit resulteren in het geheel of gedeeltelijk intrekken van de subsidie en/of het terugvorderen van reeds ontvangen voorschotbetalingen, inclusief wettelijke rente. Daarnaast staan in de beschikkingsbrief de volgende aanvullende project-specifieke verplichtingen beschreven:

- *Looptijd:* De activiteiten worden uitgevoerd volgens de aanvraag en worden uiterlijk afgerond op 31 december 2031. Activiteiten voor 1 januari 2024 en na 31 december 2031 komen niet in aanmerking voor subsidie.
- *Samenwerkingsovereenkomst:* Uiterlijk 2 april 2024 dient er een samenwerkingsovereenkomst ondertekend door alle partijen verstuurd te zijn naar RVO.
- *NHL Stenden Hogeschool:* Dient een rapport van bevindingen over de integrale kostensystematiek aan RVO te versturen voor vaststelling van het project.
- *Tata Steel:* Zolang een JLOI (Joint Letter Of Intent: maatwerkafspraken over het versneld verminderen van de CO₂-uitstoot en vermindering van de impact op de leefomgeving) niet is ondertekend wordt er geen voorschot uitbetaald aan TSN. De overige deelnemers in het project worden wel uitbetaald. Na ondertekening van de JLOI zal met terugwerkende kracht het voorschot worden verstrekt.

Tot op heden is er voldaan aan het ondertekenen van de samenwerkingsovereenkomst en het versturen van de integrale kostensystematiek.

¹¹ RVO (2024) *Handleiding subsidieregeling NGF*. Geraadpleegd op 4 oktober 2024. <https://www.rvo.nl/sites/default/files/2023-11/Handleiding%20subsieregeling%20Nationaal%20Groeifonds%202024-2025.pdf>

2.3 Charging Energy Hubs



2.3.1 Omschrijving project

Het project **Charging Energy Hubs** heeft een duur van 4 jaar, kent een begroting van ruim 70 miljoen euro en heeft 39 miljoen euro subsidie ontvangen vanuit het NGF. Het project heeft als doel om de uitrol van de oplaadinfrastructuur voor de logistieke sector te versnellen. Het project is mede opgezet vanwege de groeiende vraag van elektrische voertuigen (mede door de opkomst van zero-emissie zones in steden), en daarmee de groeiende vraag naar (krachtige) oplaadpunten. Het project sluit daarmee aan op het vinden van een oplossing voor de problematiek rondom de toenemende belasting op het elektriciteitsnetwerk.

Om een efficiënte en flexibele energiedistributie voor elektrische voertuigen te bereiken worden lokale, slimme energiesystemen ontwikkeld die de integratie van oplaadinfrastructuur en hernieuwbare energiebronnen optimaliseren. Het project heeft hiervoor 10 werkpakketten uitgezet:

- *Werkpakket 1:* Coördinatie en beheer van het project.
- *Werkpakket 2 t/m 5:* Deze werkpakketten draaien om kennisverdieping en ontwikkeling van nieuwe standaarden en technologieën rondom het gereedmaken van o.a. DC-microgrids, batterijopslag en automatische laadsystemen.
- *Werkpakketten 6 t/m 8:* In deze werkpakketten worden technologieën toegepast door middel van use-cases op verschillende plekken in Nederland, waaronder: 1) een private hub voor zero-emissie stadslogistiek; 2) een openbare hub voor commercieel vrachtverkeer, en 3) een hybride (publiek-privaat) hub voor industrieterreinen.
- *Werkpakket 9:* Kennisspreiding van inzichten en best-practices van technologieën die tijdens het project worden ontwikkeld en getest.
- *Werkpakket 10:* bevorderen van ontwikkeling van relevante kennis door de verbinding tussen belanghebbende (bedrijven, onderwijsinstellingen) in de laadinfra te versterken.

Charging Energy Hubs bestaat uit een consortium van ruim 30 partners waaronder bedrijven en kennisinstellingen zoals Heliox, Prodrive en de TU Eindhoven.

2.3.2 Toegekende subsidiebedragen en eindbegunstigden

Charging Energy Hubs heeft in totaal een budget van €75,2 miljoen, waarvan €38,9 miljoen wordt gefinancierd vanuit het NGF. Onderstaande tabel geeft de verdeling weer tussen de deelnemende partijen:

Tabel 3: Subsidiebedrag CEH per deelnemer. Bron: Verleningsbrief Charging Energy Hubs

Deelnemer	Bedrag	% van subsidie
Heliox B.V.	€ 6.743.690	17,3%
TNO	€ 4.126.844	10,6%
Technische Universiteit Eindhoven	€ 3.295.051	8,5%
Prodrive Technologies N.E. B.V.	€ 2.668.800	6,9%
TSN Houten Groen B.V.	€ 2.622.344	6,7%
Scholt Energy Control B.V.	€ 2.011.590	5,2%
DC-Systems B.V.	€ 1.612.200	4,1%
Stichting ElaadNL	€ 1.520.505	3,9%
ROCSYS B.V.	€ 1.255.800	3,2%
Maxem Energy Solutions B.V.	€ 1.252.248	3,2%
DENS Power B.V.	€ 1.123.714	2,9%
Dynniq Energy B.V.	€ 1.018.560	2,6%
Applied Micro Electronics "AME" B.V.	€ 972.602	2,5%
Van Kessel olie B.V.	€ 935.749	2,4%
Zero Emission Services B.V.	€ 866.025	2,2%
Damen Workboats B.V.	€ 739.931	1,9%
Renewable Energy Factory B.V.	€ 717.108	1,8%
Stichting Hogeschool van Arnhem en Nijmegen	€ 700.037	1,8%
Sycada Nederland BV	€ 687.664	1,8%
DAF Trucks N.V.	€ 667.604	1,7%
Shell Global Solutions International B.V.	€ 572.723	1,5%
Recoy B.V.	€ 479.253	1,2%
Firan B.V.	€ 334.895	0,9%
Fontys Hogeschool	€ 298.400	0,8%
Fluidwell B.V.	€ 268.625	0,7%
Van Berkel Logistic Investments B.V. (Van Berkel Logistics)	€ 246.190	0,6%
Damen Research Development & Innovation B.V.	€ 241.425	0,6%
KEMA B.V.	€ 214.970	0,6%
Van Berkel Investments B.V. (Van Berkel Landschap & Infra)	€ 208.782	0,5%
Prodrive Technologies B.V.	€ 172.077	0,4%

Deelnemer	Bedrag	% van subsidie
Nationaal Kennisplatform Laadinfrastructuur	€ 159.760	0,4%
FIER B.V.	€ 144.000	0,4%
XYZ Dynamics B.V. ¹²	-	-
Totaal	€ 38.879.165	100%

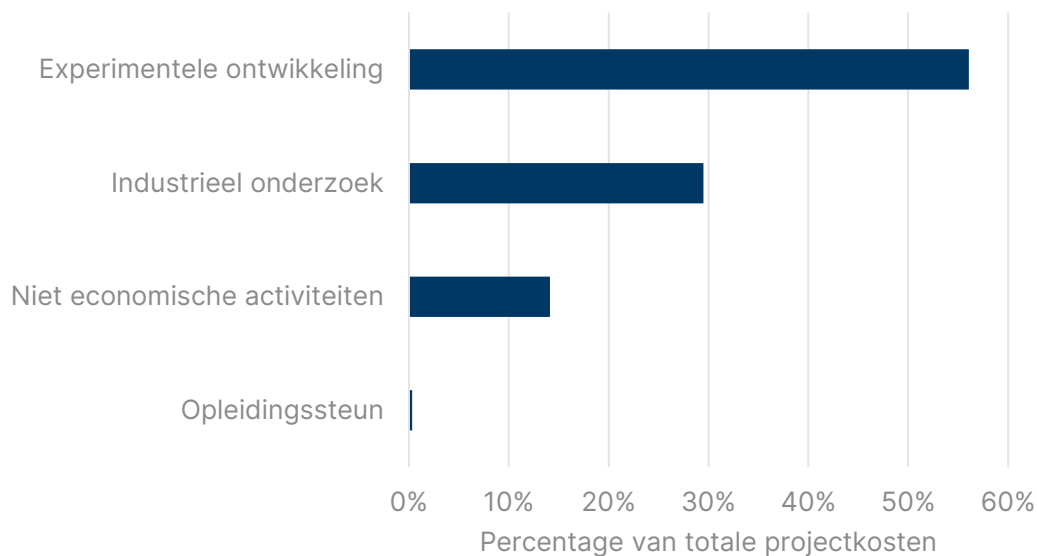
2.3.3 Voor welke activiteiten zijn de subsidie verleend?

Zoals te lezen in 2.3.1 bestaat het programma uit een groot aantal werkpakketten. Deze activiteiten worden met mede behulp van het groeifonds gefinancierd. Volgens de voorwaarden van het NGF moeten activiteiten vallen onder 1) kennisontwikkeling, en/of 2) Onderzoek, Ontwikkeling en Innovatie (O&O&I)¹³. Voor Charging Energie Hubs geldt volgens de ingediende projectbegroting dat de activiteiten voor 2% vallen onder kennisontwikkeling, en voor 98% onder O&O&I. Deze verdeling kan verklaard worden door de aard van het project, dat zich primair richt op de praktische toepassing en ontwikkeling van technologieën voor het opladen van elektrische vrachtwagens. De focus ligt daardoor minder op fundamentele kennisontwikkeling. Er geldt een afbakening voor activiteiten die binnen de subsidiabele activiteiten van het NGF vallen. Artikelen 25 t/m 29 en 31 van de AGVV zijn hierbij leidend. De totale projectkosten zijn volgens de ingediende projectbegroting als volgt verdeeld over de desbetreffende categorieën¹⁴:

¹² XYZ Dynamics B.V. heeft alleen een eigen inbreng.

¹³ RVO (2024) *Handleiding subsidieregeling NGF*. Geraadpleegd op 4 oktober 2024. <https://www.rvo.nl/sites/default/files/2023-11/Handleiding%20subsidieregeling%20Nationaal%20Groeifonds%202024-2025.pdf>

¹⁴ Deze uitsplitsing is gebaseerd op de projectbegroting die voor de verlening is opgesteld. Bij de verlening zijn er een aantal wijzigingen doorgevoerd en is de hoogte van de subsidie aangepast wat ertoe kan leiden dat de percentages per activiteit na verlening veranderd zijn. De huidige weergave geeft een te hoge inschatting van het percentage van de kosten voor industrieel onderzoek. RVO heeft bij de subsidiebeschikking sommige activiteiten anders geclassificeerd dan hoe de indieners dat hebben gedaan (bijv. als Experimentele Ontwikkeling in plaats van Industrieel onderzoek) waardoor de uiteindelijke percentages zullen afwijken van deze figuur.



Figuur 2: Type activiteiten CEH

2.3.4 Welke verplichtingen zijn verbonden aan de subsidiebeschikking?

Voor Charging Energy Hubs gelden de algemene NGF-voorwaarden waaraan moet worden voldaan, zie hiervoor de website van de RVO¹⁵. Alle deelnemende partijen zijn aan de voorgeschreven verplichtingen gebonden. Indien een partij zich niet aan de verplichtingen houdt, kan dit resulteren in het geheel of gedeeltelijk intrekken van de subsidie en/of het terugvorderen van reeds ontvangen voorschotbetalingen, inclusief wettelijke rente. Daarnaast staat in de beschikkingsbrief de volgende aanvullende project-specifieke verplichtingen beschreven:

- *Samenwerkingsovereenkomst*: Uiterlijk 1 februari 2024 dient er een samenwerkingsovereenkomst ondertekend door alle partijen verstuurd te zijn naar RVO.
- *Damen Research Development & Innovation B.V.*: Dient een rapport van bevindingen over de integrale kostensystematiek aan RVO te versturen voor vaststelling van het project

Tot op heden is er voldaan aan het ondertekenen van de samenwerkingsovereenkomst.

¹⁵ RVO (2024) *Handleiding subsidieregeling NGF*. Geraadpleegd op 4 oktober 2024. <https://www.rvo.nl/sites/default/files/2023-11/Handleiding%20subsidieregeling%20Nationaal%20Groeifonds%202024-2025.pdf>

Hoofdstuk 3

Evaluatiebegrippen en evaluatiemogelijkheden

In dit hoofdstuk oriënteren we ons op mogelijke evaluatiemethoden voor het evalueren van NGF-projecten.

3 Evaluatiebegrippen en -mogelijkheden

In dit hoofdstuk oriënteren we ons op mogelijke evaluatiemethoden voor het evalueren van NGF-projecten. Daarvoor staan we eerst stil bij enkele belangrijke begrippen en de relatie daartussen (sectie 3.1). Vervolgens gaan we nader in op diverse soorten theorie-gebaseerde evaluatie-aanpakken die we relevant achten in de NGF-context (sectie 3.2).

3.1 Bouwstenen voor evaluatiemethoden

Om het denken over geschikte evaluatiemethoden te structureren, definiëren we in deze sectie eerst enkele belangrijke begrippen die soms impliciet blijven (maar daardoor wel voor verwarring kunnen zorgen) of door elkaar gebruikt worden. Vervolgens laten we zien hoe die begrippen zich tot elkaar verhouden.¹⁶

3.1.1 Begrippen

Evaluatiemethoden

Met de term ‘evaluatiemethoden’ verwijzen we naar alles wat te maken heeft met methodologie voor het uitvoeren van evaluatieonderzoek. Dat omvat concrete analyse- en dataverzamelingstechnieken, maar ook de evaluatieaanpakken die beschrijven welke techniek op welke manier wordt ingezet. Evaluatiemethoden beschouwen we dus als een overkoepelend begrip. De bruikbaarheid van mogelijke evaluatiemethoden hangt af van de doelen van evaluatie-inspanningen en het gehanteerde perspectief op evaluatie (zie hieronder).

Evaluatiedoelen

Bij (beleids)programma’s, zeker als ze omvangrijk zijn, is er behoefte aan monitoring en evaluatie. Eén typisch doel is verantwoording: zijn de middelen goed besteed? Dergelijke **summatieve** evaluatie vraagt bijvoorbeeld om effectmeting (om ‘output-additionaliteit’ vast te stellen), al kan het ook zinvol zijn om de beleidsonderbouwing te beoordelen. Een ander doel is leren en begrijpen: via welke mechanismen werkt het programma, en voor wie? Wat kan er beter? Dit is aan de orde bij **formatieve** evaluatie. Een evaluatie hoeft zich in de praktijk niet te beperken tot één type doel: binnen dezelfde evaluatie kunnen beide typen doelen aan de orde zijn.

De twee soorten evaluatiedoelen worden in de regel geassocieerd met andere manieren om evaluatieonderzoek vorm te geven, o.a. als het gaat om wie er betrokken zijn, hoe regelmatig er data verzameld worden, welke *sample size* geschikt is, en welke vragen er gesteld worden. Er is echter maar tot op beperkte hoogte een relatie tussen

¹⁶ Veel van de inhoud van deze sectie is ontleend aan: Van der Knaap, P., Pattyn, V. & Hanemaayer, D. (2020). Beleidsevaluatie in theorie en praktijk.

evaluatie-doelen en de hieronder behandelde concrete evaluatietechnieken. Zo zien we bij verantwoordende evaluaties vaak quasi-experimentele technieken, zoals econometrie, die vooral iets zeggen over óf een effect optreedt, en hoe groot dat is. Als onderdeel van zulke analyses ontstaat er ook vaak enig inzicht in bij wie of wanneer effect zich meer of minder voordoet. Die informatie kan gebruikt worden voor formatieve doeleinden, in de zin van het verbeteren van beleid (bijvoorbeeld qua doelgroepbereik, of het beter afstellen van prikkels op eigenschappen van specifieke doelgroepen).

Evaluatiestromingen/-tradities

Er bestaan fundamenteel verschillende visies op wat de aard van evaluatie is. Een gangbaar onderscheid is het verschil tussen de positivistische stroming en de constructivistische stroming. Deze kunnen worden gezien als twee uitersten, met daar tussenin gelegen een *realist* traditie die elementen van beide visies bevat.¹⁷

Vanuit een **positivistische** kijk op evaluatie is het belangrijk om met een hoge mate van causale bewijskracht uitspraken te kunnen doen over de mate waarin uitkomsten te attribueren zijn aan de verstrekte beleidsprikkel. Het type evaluatie dat hiervoor doorgaans wordt ingezet is gebaseerd op het identificeren van *counterfactuals*, vergelijkingsbasis-sen die het mogelijk maken om het effect van het beleid te destilleren. Hierbij kan gedacht worden aan controlegroepen van organisaties, steden, landen, etc. die sterk lijken op de behandelgroep. Het streven is om zo 'zuiver' mogelijke metingen te verrichten, zodat we generaliseerbare uitspraken kunnen doen over het effect van een bepaalde interventie.

De **constructivistische** kijk op evaluatie wordt gekenmerkt door rekening te houden met verschillen in context en waarden. Of beleid succesvol is hangt af van onder meer andere beleidsmaatregelen en condities die plaats- en tijdgevoelig kunnen zijn. Bovendien is er binnen dit perspectief niet per se consensus over wanneer beleid 'goed' is; stakeholders als gebruikers of uiteenlopende beleidsmakers kunnen verschillen in hoe ze daarover oordelen, en zullen zich daar dan ook anders naar gedragen. Dit kan vervolgens weer van invloed zijn op hoe beleid uitwerkt. Diverse beleidsmakers en -uitvoerders kunnen bijvoorbeeld tegenstrijdige acties ondernemen als ze verschillen in wat ze van het beleid verwachten. Evaluatie vanuit constructivistisch perspectief beoogt dit bloot te leggen, bijvoorbeeld door *theories of change* op te stellen die expliciet maken via welke stappen beleidsinzet tot gewenste uitkomsten dienst te leiden.

De **realist** traditie is een pragmatische mengvorm. Uitgangspunt is doorgaans om met gevoel voor context, historie en waardeverschillen uitspraken te doen over wat er met beleid bereikt is. Daarin kan gebruik worden gemaakt van de *theories of change* uit de constructivistische school, maar ook van kwantitatieve methoden om na te lopen of de

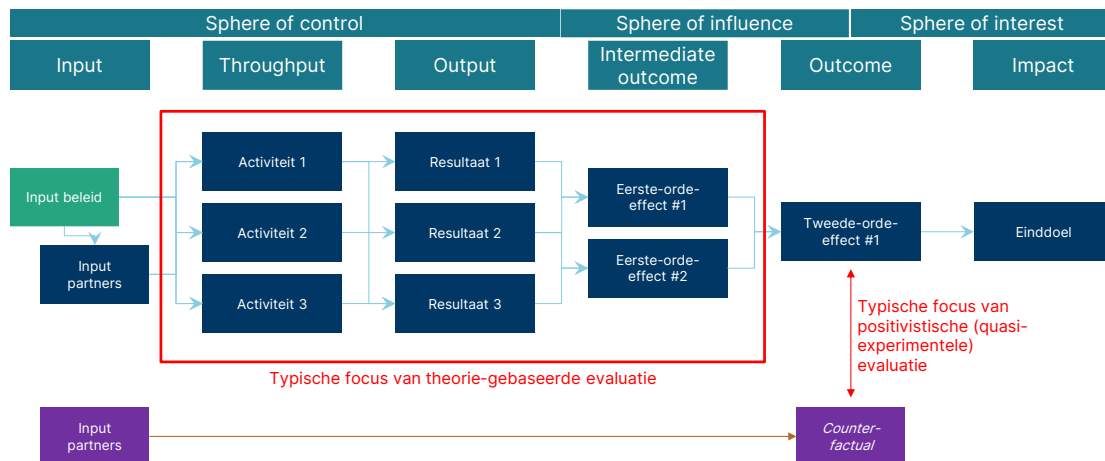
¹⁷ Arnold et al. (2018). How should we evaluate complex programmes for innovation and socio-technical transitions? Technopolis Group.

beoogde veranderingen zich ook werkelijk voordoen. Evaluaties die dit perspectief volgen kunnen zowel formatieve als summatieve doelen dienen.

Evaluatietypen

Nauw gerelateerd aan bovenstaand onderscheid is het evaluatietype dat overwogen kan worden. Zoals reeds benoemd is evaluatie vanuit positivistisch perspectief sterk **counterfactual-gebaseerd**. Des te beter de vergelijkingsbasis, des te groter de kansen om te schatten welk deel van geobserveerde ontwikkelingen te attribueren is aan het gevoerde beleid. De figuur hieronder laat zien waar de aandacht in dat soort evaluatie vooral naar uit gaat: uitkomsten aan het einde van de *Theory of Change*, die weerspiegelen waar het bij beleid uiteindelijk om te doen is. Hoe die *Theory of Change* eruit ziet maakt weinig uit (men spreekt vaak van 'black box'), zolang het mogelijk is om een goede counterfactual te identificeren of construeren. Wanneer dat lukt kan er bepaald worden óf beleid effect heeft, zonder dat dit kennis vereist over h oe dat effect tot stand komt.

De situatie is anders bij evaluaties vanuit constructivistisch (en vaak ook realist) perspectief, die in de regel **theorie-gebaseerd** zijn. Dat wil zeggen dat de *Theory of Change* een centrale rol vervult. Het opstellen daarvan kan op zichzelf al een formatief proces zijn, bijvoorbeeld doordat stakeholders hun idee en expliciet maken en op verschillen stuiten die de aanzet kunnen vormen voor herzieningen. De *Theory of Change* kan zelf ook beoordeeld worden. Tenslotte biedt het een basis om te gaan onderzoeken of de beoogde ketens van acties en gevolgen zich ook werkelijk voordoen (zie sectie 3.2).



Figuur 3: Algemene Theory of Change, met daarin de verschillende gebruikelijke foci van positivistische en theorie-gebaseerde evaluatie.

NB: in de praktijk maken beleidsevaluaties die counterfactual-gebaseerd zijn vaak ook gebruik van een beleidstheorie waarmee de werkingmechanismen (op hoofdlijnen) beschreven worden. Voor het opstellen van een goed econometrisch onderzoeksdesign is deze kennis van werkingmechanismen vaak van grote waarde, omdat [1] er meestal niet eenvoudig een perfecte controlegroep aan te wijzen is (een randomized controlled trial is vaak niet mogelijk) en [2] er binnen de experimentele groep en controlegroep gecontroleerd moet worden voor be nvloedende factoren. Zonder

voldoende kennis over de werkingsmechanismen is het moeilijk, dan niet ondoenlijk, om (alle) relevante beïnvloedende factoren in beeld te hebben en vervolgens mee te kunnen nemen in het onderzoeksontwerp.

Los van bovenstaande onderscheid tussen counterfactual-gebaseerd en theorie-gebaseerd zijn er ook andere dimensies om evaluatietypen op te categoriseren. Zo wordt er ook wel gesproken van enerzijds '**responsieve**' evaluaties met een nadruk op participatie (van beleidsbetrokkenen en andere stakeholders) en anderzijds '**reguliere**' evaluaties gericht op onafhankelijkheid en beleidsverantwoording. Een mengvorm daarvan is het mede door het Planbureau voor de Leefomgeving (PBL) en Vrije Universiteit (VU) Amsterdam ontwikkelde '**Lerend evalueren**', waarin zowel ruimte is voor co-creatie als voor het onafhankelijk oordelen over doelbereik en verbetermogelijkheden.¹⁸

Evaluatieaanpakken

Onder evaluatieaanpakken verstaan we benaderingen die de algehele opzet en uitvoering van een evaluatieonderzoek voorschrijven.

Bij counterfactual-gebaseerd onderzoek gaat het meestal om een **(quasi-)experimentele benadering** die het vergelijken van een behandelgroep en een controlegroep nabootst. Het bijbehorende proces draait dan om het verzamelen van data; het specificeren van een econometrisch model (met vaak als onderdeel het construeren van de *counterfactual*); het prepareren van de data; het uitvoeren van de analyse (het schatten van het model); en het doen van enkele tests om te kijken hoe robuust de gevonden resultaten zijn.

Voor het uitvoeren van theorie-gebaseerd evaluatieonderzoek zijn er diverse aanpakken, afkomstig uit uiteenlopende beleidsvelden. Een belangrijk onderscheid daarbij is of de *theory of change* nagelopen wordt van links naar rechts (dus van inputs naar uitkomsten), of van rechts naar links. Dat eerste is aan de orde bij **contribution analysis** en bij **process tracing**; evaluatieaanpakken voor het op systematische wijze bepalen van de bijdrage van interventies aan geobserveerde uitkomsten of impacts. Dit begint met het nalopen van de inputs, dan de throughputs, dan de outputs, enzovoorts. Onderdeel van de aanpak is om te kijken of bestaand en aanvullend bewijs consistent is met de *Theory of Change*, en om alternatieve verklaringen uit te sluiten. Bij **outcome mapping/harvesting** wordt de omgekeerde route gevolgd: evaluatoren richten zich op uitkomsten, en proberen te achterhalen welke factoren hebben bijgedragen aan het optreden van die uitkomsten. Voor die factoren wordt weer nagegaan waarom ze zich hebben voorgedaan. Op die manier komt het onderzoek uiteindelijk al dan niet uit bij beleidsinputs. Outcome mapping/harvesting wordt doorgaans niet ingezet wanneer er een duidelijke

¹⁸ 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. Amsterdam: Athena Instituut, VU.

Theory of Change is opgesteld, maar juist wanneer er ook nog verkend dient te worden wat voor (gedrags)veranderingen er hebben opgetreden.

Analysetechnieken en data(verzameling)

Op het meest concrete niveau van het hier behandelde evaluatiejargon vinden we uiteindelijk de analysetechnieken en dataverzamelingsmethoden.

De technieken die geschikt zijn voor de uitvoering van quasi-experimentele evaluatie-aanpakken zijn terug te vinden in het befaamde rapport “Durf te meten” van de Expertwerkgroep Effectmeting (Commissie Theeuwes, 2012). Dit rapport richt zich op de specifieke context van innovatiebeleid, maar bouwt voort op decennia aan economisch onderzoek. Vanuit dat perspectief vormen natuurlijke en sociale experimenten de ‘gouden standaard’ voor evaluatie, en is het zaak om dergelijke experimenten zoveel mogelijk te benaderen via statistische (econometrische) technieken – vandaar de naam quasi-experimenteel. Voor het zo zuiver mogelijk meten van beleidseffect is het zaak dat verschillen in prestaties zo min mogelijk samenhangen met eigenschappen van de groepen die met elkaar vergeleken worden, inclusief mogelijke biases als gevolg van zelfselectie. Voorbeelden van geschikte analysetechnieken zijn *regression discontinuity* (waarbij effecten worden geschat door te kijken of bijvoorbeeld drempelwaardes in beleid samenhangen met verschillen in prestaties); *difference-in-difference* (een vergelijking tussen twee groepen, door de tijd heen), en *propensity score matching* (een manier om uit veel observaties een controlegroep en behandelgroep samen te stellen die zo veel mogelijk op elkaar lijken). Sommige van die regressietechnieken kunnen ook in combinatie met elkaar gebruikt worden. Los van de familie van econometrische technieken zijn er ook andere technieken die dienen om systematisch te bepalen of beleid van invloed is geweest op geobserveerde uitkomsten. Een voorbeeld is *qualitative comparative analysis* (QCA); een techniek die zich leent om te kijken welke combinatie van factoren (waaronder mogelijk ook beleidsinterventie) variatie in succes bepaalt.

Voor theorie-gebaseerde evaluatieaanpakken uit de constructivist en realist traditie is het iets lastiger om concrete analysetechnieken te benoemen. Vaak bestaan die aanpakken uit een reeds stappen, die zowel gaan over het opstellen van een *Theory of Change* als het toetsen daarvan. Dat toetsen kan aan de hand van literatuur (is het aan nemelijk dat X tot Y leidt?), maar ook empirisch. De essentie van die empirische analyse is dan om te inspecteren of de realiteit consistent is met wat de *Theory of Change* beschrijft. Sommige evaluatieaanpakken onderscheiden een reeks van concrete tests, bijvoorbeeld ook om alternatieve verklaringen voor geobserveerde ontwikkelingen uit te sluiten. Het uitvoeren van die tests kan middels allerlei technieken, zowel kwalitatief (bijv. case studies) als kwantitatief (bijv. benchmarking).

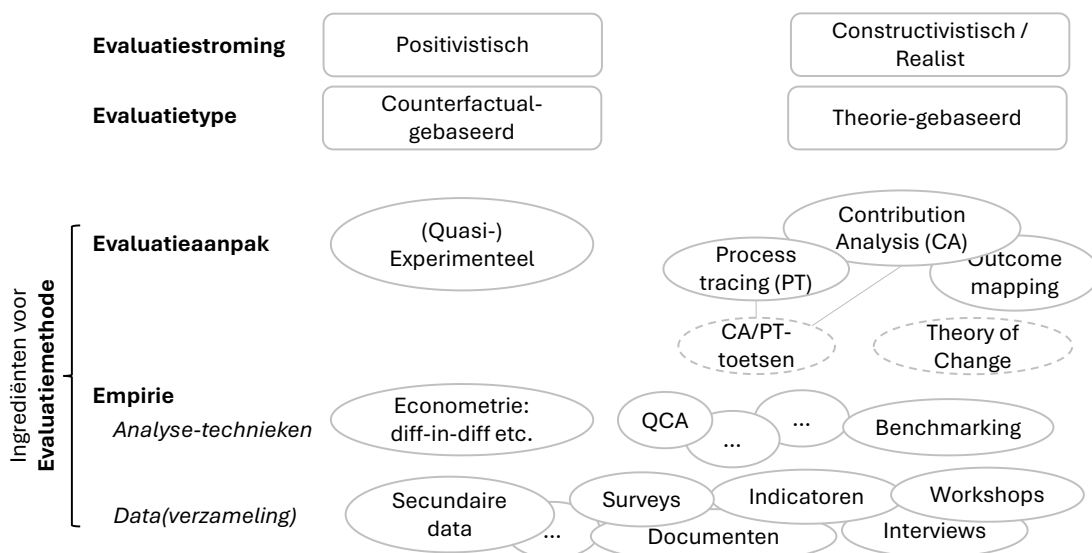
Het toepassen van empirische analysetechnieken vergt data. Primaire data kunnen verzameld worden via bijvoorbeeld interviews en workshops. Daarnaast zijn er ook bronnen die secundaire data bevatten, zoals registers met patenten en publicaties of de CBS-microdatabestanden. Er is een relatief nauwe samenhang tussen sommige

analysetechnieken en datatypen (econometrie is gebaat bij grote databestanden met veel variabelen), maar er zijn ook datatypen die wat lastiger te plaatsen zijn. Zo kunnen surveys en indicatoren-overzichten voor veel soorten analyses gebruikt worden, en zijn er dankzij *large language models* steeds meer mogelijkheden om kwantitatieve analyses uit te voeren op (ongestructureerde) teksten.

De belangrijkste *take-away* uit bovenstaande begrippenlijst is dat het niet eenvoudig is om een lijstje evaluatiemethoden op te stellen, en dan voor iedere methode te bepalen hoe geschikt die is voor toepassing in de context van NGF-projecten. Een begrip als *contribution analysis* (evaluatieaanpak) is van een hele andere orde dan een econometrische regressietechniek (analysetechniek) of een survey (dataverzamelmethode). Een evaluatiemethode omvat die begripsniveaus, waarbij er ook allerlei combinaties gemaakt kunnen worden. De volgende sectie gaat hier verder op in.

3.1.2 Relatie tussen de begrippen

Bij het behandelen van de begrippen, en de verschillende manieren waarop die invulling kunnen krijgen, is steeds al kort aangestipt dat er relaties bestaan tussen evaluatiedoe- len, - stromingen/-typen, -aanpakken en analysetechnieken. Figuur 4 hieronder brengt dit visueel in kaart. Theory of Change wordt soms ook als evaluatieaanpak beschouwd, maar hier zien we theories of change meer als de structurerende onderdelen van (theo- rie-gebaseerde) evaluatieaanpakken die richting geven aan het verzamelen en analyseren van empirische data. Vandaar dat we deze tussen evaluatieaanpak en empi- rie in gepositioneerd hebben. Hetzelfde geldt voor de diverse soorten toetsen om bij contributie-analyse of process tracing uitspraken te doen over de causaliteit in waarge- nomen verbanden; zie ook sectie 3.2.1.



Figuur 4: Evaluatiebegrippen, mogelijke invullingen, en hun positionering ten opzichte van elkaar.

Een belangrijke kanttekening bij de figuur is dat deze wellicht wat misleidend is in de manier waarop het twee perspectieven met elkaar contrasteert. **In de praktijk kunnen deze perspectieven met elkaar gecombineerd worden.** Omwille van de duidelijkheid is in de begripsbeschrijving hierboven beschreven wat gangbare relaties zijn, maar die relaties zijn niet inherent of exclusief. Zo benoemden we al dat positivistisch counterfactual-gebaseerd evaluatieonderzoek vaak een summatief beleidsdoel dient, maar ook inzichten kan genereren die voor formatieve evaluatie relevant kunnen zijn. Op die manier zijn er veel dwarsverbanden te vinden. Sommige van de theorie-gebaseerde evaluatieaanpakken zijn bijvoorbeeld geschikt voor zowel het vaststellen en verbeteren van *theories of change* (en dus het vergroten van beleidseffect) als voor het toetsen van de invloed van beleid ten opzichte van andere factoren. En ook het intuïtieve onderscheid tussen een kwalitatief en een kwantitatief paradigma is niet terecht; bij het toetsen van *theories of change* kan zeker ook gebruik worden gemaakt van indicatoren-overzichten (bijv. van NGF-consortia) en statistische analyses.

Behalve dat er tussen de begrippen weinig één-op-één relaties zijn, is er ook nog het punt dat **veel van de besproken aanpakken en technieken op allerlei manieren te combineren zijn.** Dit is een centrale boodschap in o.a. het rapport “Durf te leren, ga door met meten”¹⁹, waarin tien jaar na publicatie van het “Durf te meten”-rapport is gekeken naar mogelijkheden om systeembeleid (en transitiebeleid) te evalueren.

Voor integrale en dynamische beleidsprogramma's, waarbij uiteenlopende en evoluerende beleidsacties worden gericht op doelgroepen die verschillend zijn, maar wel onderdeel vormen van hetzelfde systeem, is het nauwelijks mogelijk om op programma-niveau quasi-experimenteel onderzoek te doen conform de “Durf te Meten”-aanbevelingen. Het voorgestelde alternatief is om bij evaluatie van dit soort beleid te beginnen met het helder krijgen van de *Theory of Change* (“durf te leren”), wat ingewikkeld kan zijn op het moment dat beleidsbetrokkenen daar hele andere beelden bij blijken te hebben. Als het lukt om de verandermechanismen te specificeren, wordt het mogelijk om stap voor stap na te gaan of die mechanismen zich ook werkelijk voordoen. Dat kan met zowel kwalitatieve als kwantitatieve analysetechnieken (“ga door met meten”). Deze manier van werken is inherent aan evaluatieaanpakken als *contribution analysis* en *process tracing*, en ook voor counterfactual-gebaseerd onderzoek geldt dat een scherpere *Theory of Change* meer mogelijkheden biedt om betrouwbare schattingen te doen.

3.2 Verdieping van kansrijke evaluatieaanpakken

Bovenstaande logica, ten aanzien van het **combineren van theorie-gebaseerde en counterfactual-gebaseerde evaluatie** (die dus niet eens helemaal los van elkaar te zien zijn), is ook relevant als het gaat om de evalueerbaarheid van NGF-projecten. Zeker de

¹⁹ ter Weel, B., Janssen, M., Bijlsma, M., & De Boer, P. J. (2022). *Durf te leren, ga door met meten: Op zoek naar kaders en methoden voor de evaluatie van systeem- en transitiebeleid.*

projecten in de OOI-pijler zijn inherent systemisch van aard, aangezien de grondslag van het NGF-beleid daar wordt gevormd door de notie van onderzoeks- en innovatie-ecosystemen.²⁰ Dat betekent dat NGF-projecten in de regel meerdere actielijnen omvatten, die er samen op gericht zijn om onderzoek, innovatie en valorisatie te versterken in het ecosysteem rondom een bepaald thema. De consortia die NGF-projecten indienden zijn opgeroepen om dat ecosysteem zo goed mogelijk te mobiliseren en te betrekken bij de beoogde actielijnen. Als gevolg is het lastig om dat ecosysteem, of de betrokken partijen, af te zetten tegen een controlegroep die daar sterk op lijkt. De verwachting is immers dat partijen voor wie het NGF-project relevant is direct (als deelnemer in activiteiten die bekostigd zijn uit het NGF-project) of indirect (als gebruiker van nieuwe kennis, opgeleid talent, etc.) voordeel ondervinden van de systeemversterking. Partijen voor wie dat niet geldt zijn waarschijnlijk niet goed vergelijkbaar, omdat ze aan hele andere (of helemaal geen) innovaties werken.

Een denkbare optie is nog om, bij wijze van benchmark, naar prestaties in vergelijkbare ecosystemen in het buitenland te kijken. Voor een relevante vergelijkende analyse is het dan ook zaak om te achterhalen hoe de beleidssteun in die andere landen eruitzag, en hoe het ecosysteem zich (al dan niet dankzij die steun) aan het ontwikkelen was. Als hier een duidelijk beeld van te maken is, kan een vergelijking waardevol zijn. Zelfs als er alleen maar ordegrottes bekend zijn van prestaties en verleende beleidssteun, kan een benchmark helpen om beter te begrijpen wat de invloed van NGF-steun is geweest. Doen andere landen het met minder steun even goed, dan is dat een indicatie dat NGF-steun mogelijk niet effectief was. Voorzichtigheid is geboden bij dit soort interpretaties, aangezien geobserveerde patronen ook kunnen samenhangen met andere factoren (zoals de volwassenheid van een ecosysteem). Vandaar dat dit soort benchmarks het best zijn in te zetten als onderdeel van een onderzoek dat uit meerdere analyses bestaat; bijvoorbeeld op basis van de verderop in deze sectie besproken evaluatieaanpakken.

Bovenstaande redeneringen gelden in grote mate ook voor de kennispijler van het NGF, waarbij de koppeling tussen beleidssteun en beoogde maatschappelijke impacts gescheiden wordt door een lange weg, zowel in termen van aantallen causale stappen als jaren. Die kennispijler richt zich namelijk op onderwijs en vaardigheden. Inputs, throughputs (activiteiten) en outputs (resultaten) zouden goed gemeten kunnen worden, maar voor effectmeting op het niveau van uitkomsten en impacts is het lastig om een betrouwbare vergelijking te doen. Het causale mechanisme is hier dat investeringen in onderwijs en vaardigheden moeten resulteren in beter / relevanter opgeleide scholieren en studenten (en/of werkenden) die vervolgens in hun professionele carrière meer in staat zijn om bij te dragen aan nieuwe producten, processen, werkwijzen, etc. (die uiteindelijk weer doorwerken in meer welvaart, in nauwe of brede zin). Dit is in theorie te onderzoeken door, over een groot aantal jaren, prestatie-ontwikkelingen van

²⁰ EZK & OCW (2020). Kamerbrief over Kabinetsstrategie 'Versterken van onderzoeks- en innovatie-ecosystemen'. Dialogic (2020). Achtergrondstudie Onderzoeks- en innovatie-ecosystemen in Nederland.

organisaties te relateren aan de mate waarin ze personeel hebben dat is blootgesteld aan activiteiten uit de kennispijler van NGF. Als die activiteiten hele cohorten of vakgebieden beslaan is het lastig om een goed contrast te vinden, te meer omdat er ook (zelf)selectie-effecten in het spel kunnen zijn als het gaat om waar scholieren en studenten komen te werken.

Op het moment dat er geen counterfactual bestaat op niveau van de NGF-projecten als zodanig, en/of als het vele jaren duurt voordat uitkomsten zich manifesteren, ligt het voor de hand om de aandacht te richten op het in kaart brengen en nalopen van de Theory of Change (de ketens van acties en gevolgen). Hieronder bespreken we enkele theorie-gebaseerde evaluatieaanpakken in meer detail, met daarbij aandacht voor manieren om bij het nalopen van de veranderketens zo betrouwbaar mogelijke (eventueel kwantitatief en counterfactual-gebaseerd) uitspraken te doen over causaliteit en attributie.

3.2.1 Contribution analysis

Contribution analysis (contributieanalyse) helpt om duidelijkheid te krijgen over de bijdrage die een (beleids)interventie heeft geleverd aan geobserveerde ontwikkelingen. De nadruk ligt niet op het vaststellen van de omvang van een beleidseffect, maar op het onderzoeken van de interne en externe factoren, processen en mechanismen die bepalen hoe de interventie een bijdrage levert aan wenselijke en onwenselijke gevolgen. Een uitgebreidere Nederlandstalige beschrijving, inclusief achtergronden, toepassingen en beperkingen, is te vinden in een recent gepubliceerd artikel (Leeuw, 2024).²¹ Onderstaande tekst is sterk gebaseerd op dit artikel en het bronmateriaal waar het naar verwijst, inclusief het werk van John Mayne, de grondlegger van contribution analysis.

In essentie is contributie-analyse een aanpak om op systematische wijze toe te werken naar 'contributie-claims' over de causale invloeden van een (beleids)interventie. De aanpak neemt een Theory of Change als uitgangspunt. Vervolgens is het zaak om voor alle elementen in die Theory of Change empirische gegevens te verzamelen, en die zodanig te analyseren dat er een beeld ontstaat van hoe waarschijnlijk het is dat de interventie van invloed is geweest op waargenomen uitkomsten. Voor het doen van uitspraken over causaliteit zijn er vier condities waaraan voldaan moet worden (Befani & Mayne, 2014, p.21):²²

1. **'Plausibility'** ('Aannemelijkheid'). De interventie moet zijn gebaseerd op een onderbouwde Theory of Change. Die keten van resultaten en de aannames (over

²¹ Leeuw, F. (2024). Contributie-analyse: achtergronden, methode, toepassingen en problemen/beperkingen. In: Beleidsonderzoek Online 2024. Themareeks: Methoden en benaderingen in de beleidsevaluatie, D. Hanemaayer, J. Mevissen, V. Pattyn (red.)

²² Befani & Mayne (2014): Process Tracing and Contribution Analysis: A Combined Approach to Generative Causal Inference for Impact Evaluation

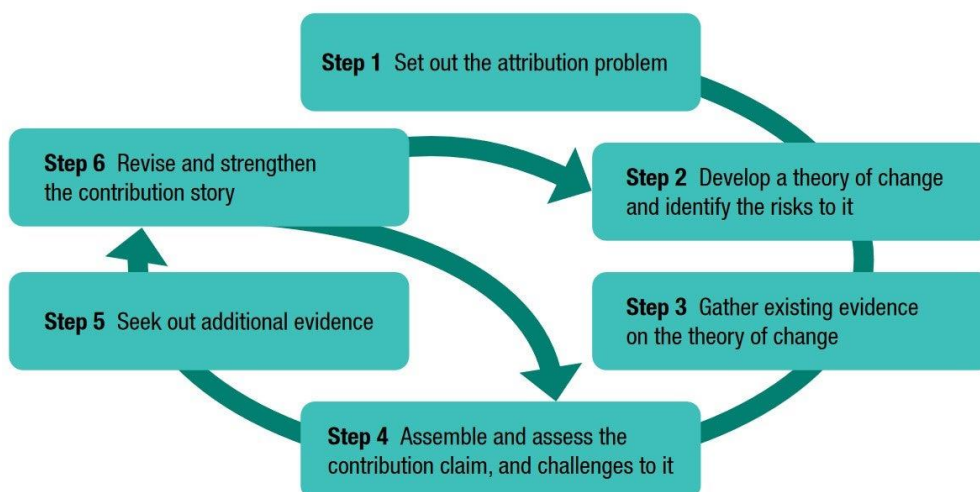
waarom de interventie naar verwachting zal werken) is plausibel, gedegen, geïnformeerd door bestaand onderzoek en literatuur, en wordt ondersteund door de belangrijkste belanghebbenden.

2. **'Fidelity'** ('Getrouwheid'). De activiteiten van de interventie zijn uitgevoerd zoals uiteengezet in de Theory of Change.
3. **'Verification of the Theory of Change'** ('Verificatie van de Theory of Change'). De Theory of Change wordt geverifieerd door bewijs: de keten van verwachte resultaten heeft plaatsgevonden en de causale aannames zijn valide.
4. **'Other influencing factors'** ('Afweging tegen andere beïnvloedende factoren'). Context en andere factoren die van invloed zijn op de interventie worden geïdentificeerd en gewogen. Op basis daarvan wordt aangetoond dat ze geen significante bijdrage hebben geleverd. Als die factoren wel van invloed waren wordt hun relatieve bijdrage erkend en opgenomen in de Theory of Change.

Dit vierde criterium zorgt er onder meer voor dat een eventuele 'positieve bias' vanuit de opstellers van de Theory of Change niet (onbewust) in de oordelen binnen het onderzoek overgenomen wordt.

Hoewel beschrijvingen van contributie-analyse doorgaans niet in letterlijke zin ingaan op falsificatie, is dat principe geborgd door de vierde conditie en door de iteraties in het te doorlopen stappenplan (zie hieronder). Samen vergen zij namelijk dat evaluatoren kritisch kijken naar de vraag of het gevonden bewijs, per beleidstheorie-element, wel echt doorslaggevend is. Als bewijzen voor gerelateerde elementen in de beleidstheorie zwak of tegenstrijdig zijn, is dat een goede reden om extra kritisch te kijken en eventueel aanvullend onderzoek te doen. Op die manier kan worden voorkomen dat evaluatoren naar gewenste resultaten toe redeneren.

Het uitvoeren van contributie-analyse, en daarmee ook het onderzoeken van bovenstaande vier condities, verloopt via het volgen van een aantal stappen. Deze zijn in Figuur 5 gevisualiseerd. We bespreken ze achtereenvolgens.



*Figuur 5: Visualisatie van de stappen in contributie-analyse.*²³

- **Stap 1: Beschrijf het attributie-probleem.** Voor welke gevolgen dienen de oorzaken te worden bepaald, en wat is de context waarin dit zich afspeelt? Bij NGF-projecten gaat het erom om ontwikkelingen op (tussen)doelen te herleiden naar de inputs vanuit het NGF, of die mogelijk gemaakt zijn door het NGF.
- **Stap 2a: Ontwikkel een *Theory of Change*.** De Theory of Change beschrijft via welke mechanismen (causale paden) inspanningen en middelen naar verwachting leiden tot lagere en hogere orde uitkomsten. Vaak wordt dit gepresenteerd in de vorm van een figuur. Dit kan een logisch raamwerk zijn (met inputs-throughputs-outputs-outcomes-impacts en de schakels daartussen; zie Figuur 3), maar als er veel dwarsverbanden zijn is bijvoorbeeld een netwerkvisualisatie ook een optie. Als er belangrijke externe factoren zijn dienen die ook meegenomen te worden. Er zijn verschillende manieren waarop de Theory of Change tot stand kan komen, en wie daar bij betrokken zijn. Zo kunnen beleidsbetrokkenen (bijv. coördinatoren van NGF-projecten) dit zelf doen; ze kunnen het overlaten aan externe evaluatoren; of het gebeurt in gezamenlijkheid.
- **Stap 2b: Bepaal met welke risico's de *Theory of Change* gepaard gaat, en wat de kwaliteit daarvan is.** Vaak is er niet zomaar consensus over de doelen en verandermechanismen die daar bij horen. Diverse stakeholders kunnen daar uiteenlopende beelden bij hebben. Wanneer er geen eenduidigheid is, is dat zeer bepalend voor hoe de rest van een evaluatie eruitziet. Evaluatoren lopen het risico dat ze uitspraken doen over doelen die door niet door iedereen erkend worden. Behalve dat een Theory of Change omstreden kan zijn, kan ook de bruikbaarheid ervan gelimiteerd zijn. Dit is bijvoorbeeld aan de orde als ze heel algemeen blijft op het definiëren van (tussen)doelen of op het beschrijven van de mechanismen. Een Theory of Change kan ook in meerdere opzichten

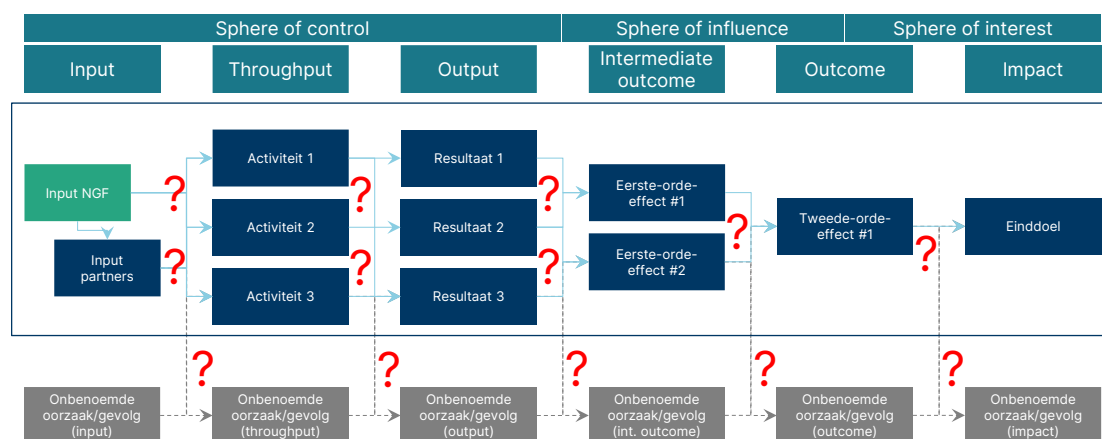
²³ Better Evaluation, o.b.v. Apgar et al., 2017

onvolledig zijn; denk aan het ontbreken van tussendoelen (waardoor onduidelijk blijft hoe acties verband houden met uitkomsten) of van belangrijke externe factoren.

- **Stap 3: Verzamel bewijsmateriaal voor het toetsen van de Theory of Change.** Dit bewijsmateriaal kan afkomstig zijn uit bestaande studies, en/of verkregen worden via aanvullende dataverzameling (via interviews, raadplegen van databases met relevante indicatoren, etcetera). De empirische inzichten dienen gekoppeld te worden aan de elementen uit de Theory of Change waar ze iets over kunnen zeggen (resultaten, schakels daartussen, andere factoren en verklaringen). Uiteraard helpt het enorm als er voor in ieder geval de resultaten in de Theory of Change indicatoren zijn opgesteld waarop gemonitord wordt.
- **Stap 4: Stel de “contributie-claims” samen.** Die claims drukken uit hoe aannemelijk het is dat de interventie een invloed heeft gehad op geobserveerde uitkomsten. Dit volgt uit het beoordelen van het empirisch bewijsmateriaal dat in stap 3 verzameld en geordend is. Voor uitspraken over causale attributie is het niet voldoende om enkel te inspecteren of waargenomen ontwikkelingen *consistent* zijn met de Theory of Change; de contributie-claim is sterker op het moment dat alternatieve verklaringen uitgesloten kunnen worden. Mogelijk zijn die alternatieve verklaringen al geïdentificeerd bij het opstellen van de Theory of Change, maar ze kunnen ook in beeld komen bij deze stap. Bij het interpreteren van de verzamelde data moet immers steeds gevraagd worden: “zeggen deze indicaties dat de voorgaande stap in de Theory of Change cruciaal was voor een bepaald resultaat, of zou er ook een andere oorzaak (*zoals zelfselectie of een confounding variable*) kunnen zijn?”. De uitdaging is om vervolgens die alternatieve verklaringen te ontkrachten. Soms vraagt dit aanvullende data-verzameling en -analyse; zie stap 5.
- **Stap 5: Zoek aanvullend bewijs.** Om uitspraken over causale relaties aannemelijk te maken kan het nodig zijn om extra onderzoek te verrichten. Zeker op dit punt kan het behulpzaam zijn om in termen van counterfactuals te denken; zijn er manieren om met extra data vast te stellen dat een causale relatie echt is terug te leiden op het beleid, of gevolgen van het beleid, en dat die relatie dus niet bestaat voor soortgelijke groepen, contexten, etc. waarin het beleid geen rol speelde? Deze vraag wordt dus niet gesteld op het niveau van de hele Theory of Change, maar op de afzonderlijke onderdelen daarvan (in de ‘black box’; zie Figuur 3). Om de causale bewijskracht te versterken kan het helpen om gebruik te maken van de toetsen uit process tracing; zie sectie 3.2.2.
- **Stap 6: Herzie en verbeter de ‘contributieclaims’ (uit stap 4) en het ‘contributieverhaal’ dat daarmee verteld kan worden.** Door bij stap 5 in detail te kijken naar de zwakste of meest onderbelichte onderdelen van de Theory of Change, ontstaan er wellicht nieuwe inzichten. Die dienen geïnterpreteerd te worden, zodat er een duidelijker beeld ontstaat over de uiteindelijke bijdrage van het beleid. Die heeft niet de vorm van een maat, maar van een onderbouwd verhaal over

welke causale paden gevalideerd zijn en waarom (niet). Hierin krijgen de bevindingen t.a.v. externe factoren en alternatieve verklaringen dus hun plek. Het is mogelijk dat sommige veranderpaden wel zijn opgetreden, en andere niet. Ook kan het zijn dat een pad zich maar tot op een zekere uitkomst in de Theory of Change heeft voorgedaan, en dat de veranderingen vanaf daar gestagneerd zijn of puur op rekening van andere invloeden komen. Dergelijke inzichten zijn relevant in zowel summatief opzicht (wat zijn de gevolgen van beleid?) als formatief opzicht (waar lopen de veranderpaden vast?).²⁴

Voortbordurend op Figuur 3 vat Figuur 6 hieronder de gedachte achter contributie-analyse nog eens samen.



Figuur 6: Bij contribution analysis wordt van links naar rechts, voor iedere stap in de Theory of Change, bekeken of de beoogde ontwikkeling zich voordoet en of dat door de voorgaande stap komt.

Voor iedere element in de Theory of Change dient bepaald te worden of dat resultaat zich voordoet, en of dat komt door de voorgaande elementen (oftewel: levert die schakel werkelijk de bijdrage?). Daarbij is het zaak om ook onbenoemde oorzaken en gevolgen te identificeren, om zo een accurater beeld te construeren van de relatieve invloed van het beleid.

In dit geval bestaat dat beleid uit de input die verstrekt is vanuit het NGF, in termen van middelen, capaciteit en aandacht. De eerste mogelijke werking die daarvan uit gaat, en die dus onderzocht moet worden bij het doorlopen van de Theory of Change, is het aantrekken van inputs vanuit de partners die deelnemen in NGF-consortia (ook wel bekend als “input-additionaliteit”). Vervolgens is het zaak om na te lopen of dit geleid heeft tot extra activiteiten (“throughput-additionaliteit”); of die geleid hebben tot extra resultaten (“output-additionaliteit”), etcetera.

²⁴ Indien de evaluatie leidt tot het herzien van de contributieclaims en het aanpassen van de Theory of Change, kan in een eventuele vervolgevaluatie de aangescherpte Theory of Change als uitgangspunt gehanteerd worden.

De betrouwbaarheid van het uiteindelijke contributieverhaal volgt uit de kwaliteit van de analyses voor de afzonderlijke elementen uit de Theory of Change. Uit die analyses samen volgt een beeld. Daarbij kan het zijn dat voor sommige elementen niet hard is uit te sluiten dat externe factoren ook van invloed waren. Als veel deel-analyses dezelfde kant op wijzen maakt dat het wel aannemelijker dat beleid een bijdrage leverde aan die ontwikkelingen, al kan het nog altijd zo zijn dat alle geobserveerde veranderingen zijn toe te schrijven aan een belangrijke externe ontwikkeling. Om die reden is het aan te bevelen om zo scherp mogelijk te toetsen op alternatieve verklaringen. Process tracing kan daarbij een uitkomst bieden. Hieronder gaan we daar verder op in.

3.2.2 Process tracing

Process tracing is een evaluatieaanpak die een andere herkomst heeft dan contribution analysis, maar qua opzet veel overeenkomsten vertoont. Kenmerkend is dat men bij process tracing onderzoekt welke combinatie van factoren noodzakelijk en voldoende is voor het zich voordoen van een causaal mechanisme. In dat opzicht vertoont het overeenkomsten met QCA, waarin eveneens gebruik wordt gemaakt van Bayesiaanse logica (probabilistische statistiek, waarbij de kans op een uitkomst wordt geschat door naar een aantal voorwaarden te kijken). Een verschil is dat QCA niet per se theorie-gebaseerd is, waardoor verklarende factoren ook buiten de Theory of Change gezocht kunnen worden. Process tracing is juist wel sterk gebonden aan de elementen die in een Theory of Change zijn opgenomen.

Net als bij contribution analysis is het doel om zo betrouwbaar mogelijke uitspraken te doen over de invloed van beleid, op basis van empirische indicaties dat bepaalde resultaten en schakels zich voordoen. Het empirische bewijs is sterker als er voor achtereenvolgende elementen in de Theory of Change indicaties zijn dat de verwachte ontwikkelingen zich voordoen, zonder dat dit volledig is toe te schrijven aan andere factoren. Het voornaamste verschil met contribution analysis is dat process tracing (PT) een viertal concrete toetsen kent voor causale gevolgtrekking. Deze zijn getoond in onderstaande tabel.^{25,26}

Tabel 4: Process tracing tests

		Is het bewijs voldoende voor de causale relatie?	
		Nee	Ja
Is het bewijs noodzakelijk voor de causale relatie?	Nee	<i>Straw-in-the-wind</i>	<i>Smoking Gun</i>
	Ja	<i>Hoop</i>	<i>Doubly Decisive</i>

²⁵ Collier D. Understanding Process Tracing. *PS: Political Science & Politics*. 2011.

²⁶ <https://www.betterevaluation.org/methods-approaches/methods/process-tracing>

- **Straw-in-de-wind:** Dit soort test biedt slechts zwakke ondersteuning voor een verklaring, omdat het gevonden bewijs ook bij andere verklaringen kan passen. Als er bijvoorbeeld meer investeringen gaan naar een nieuwe technologie kan dat komen door NGF-investeringen, maar misschien was de technologie al in opkomst (en is dat waarom er een NGF-project op is gevormd).
- **Hoop:** Deze test stelt noodzakelijke criteria vast waaraan een causale verklaring moet voldoen; de naam verwijst naar een 'hoepel' waar het bewijs door moet kunnen om aannemelijk te blijven. Het niet slagen voor een hoop-test elimineert een mogelijke causale verklaring. Wél slagen voor een hoop-test biedt ondersteuning voor een mogelijke causale verklaring, maar bewijst deze niet. Een hoop-test kan bijvoorbeeld de volgorde der dingen betreffen - een mogelijke oorzaak moet in de tijd zijn voorafgegaan aan de uitkomst.
- **Smoking gun:** De *smoking gun*-test is te doorstaan als er voldoende bewijs is om een causale verklaring te ondersteunen. De naam verwijst naar het vinden van een moordverdachte met een recent afgevuurd pistool. Het slagen voor deze test biedt sterke ondersteuning voor een causale verklaring, en verzwakt alternatieve verklaringen zonder ze definitief uit te sluiten. Het niet slagen voor een smoking gun-test elimineert een mogelijke causale verklaring niet, maar verzwakt deze wel.
- **Double decisive:** Het doorstaan van de *doubly decisive*-test levert sterk bewijs dat zowel een bepaalde causale verklaring bevestigt, als andere verklaringen uitsluit.

De *Straw-in-the-wind*-toets lijkt dus het meest op correlatie zonder bewijs over de aanwezigheid of richting van een causale relatie, terwijl *Doubly decisive* de strengste toets is. Die toets correspondeert grofweg met hoe er in quasi-experimenteel onderzoek wordt geanalyseerd. Een praktisch verschil is dat in quasi-experimenteel onderzoek vaak de kracht van grote getallen wordt benut, terwijl men bij PT erkent dat er ook met een klein aantal observaties sterk bewijs voor een causale relatie kan zijn.²² Als een NGF-project bijvoorbeeld een nieuwe technologie ontwikkelt die gepatenteerd wordt, en vervolgens zijn er partijen die een licentie nemen op dat patent, dan is dat een relatief sterke indicatie dat er sprake is van diffusie. Zo'n geobserveerde directe relatie zegt meer dan wanneer er met proxy-variabelen (die slechts bij benadering iets zeggen over het soort technologie dat bedrijven gebruiken) een verschil tussen een behandelgroep en een controlegroep wordt geschat.

In de regel is het verstandig om een zo streng mogelijke toets na te streven. Een verdedigbare vuistregel is echter "Hoe plausibeler de ToC, hoe minder streng de benodigde test hoeft te zijn". Dit komt doordat de aannemelijkheid van de hypothese op voorhand al groter is en alternatieve hypothesen al minder aannemelijk zijn. Als de plausibiliteit extreem laag is, bijv. doordat er niet eens een ToC is, moet de toets juist heel streng zijn. De toets zou zo streng moeten zijn dat je zelfs zonder enige logische onderbouwing van

hoé het zou moeten werken kunt vaststellen of er een causale relatie is. Als de plausibiliteit extreem hoog is, bijv. doordat er een gedetailleerd en scherp uitgewerkte ToC is inclusief aannames en externe factoren buiten de directe scope van het voorstel, kan de toets vermoedelijk minder streng zijn. Het monitoren van alle elementen en schakels in de Theory of Change, zonder verdere zware toets kan voldoende zijn om als evaluatoren causaliteit voldoende te onderbouwen.

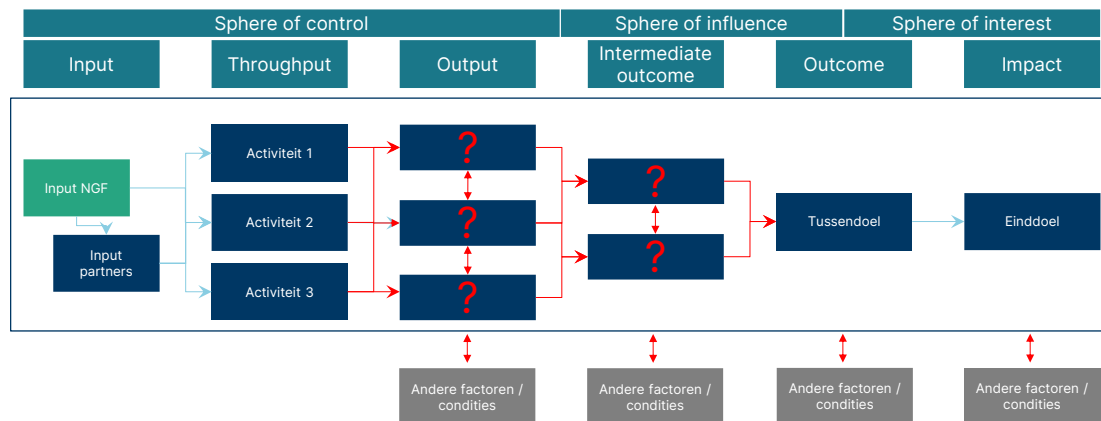
Bij het geven van een invulling aan (één van) de vier PT-testen kunnen verschillende analysetechnieken en dataverzamelingstechnieken ingezet worden. Ter illustratie geven we hier enkele voorbeelden:

- Benchmarken kan een manier zijn om empirisch invulling te geven aan een toets (bijv. “Nederland dient na 5 jaar koploper te zijn in de EU”);
- Een econometrische analyse zoals difference-in-difference-regressie *kán* een techniek zijn om aan een *doubly decisive test* te voldoen (bijv. “ontvangers van ondernemersadvies krijgen hun product sneller op de markt”);
- Een online enquête kan ingezet worden om te achterhalen hoe mensen denken over een bepaalde technologie. Als een doel van het beleid is om dat beeld van mensen te veranderen kan deze enquête-data gebruikt worden om een process tracing test te toetsen.

Befani & Mayne hebben betoogd dat de PT-toetsen een mooie aanvulling zijn op CA, omdat het mogelijk wordt om met meer vertrouwen iets te kunnen zeggen over de impact van een interventie. Binnen CA wordt er geen expliciet onderscheid gemaakt naar de verschillende soorten indicaties dat er sprake kan zijn van een causale relatie. Door per relatie een toets te bedenken die zo dicht mogelijk bij ‘doubly decisive’ komt, wordt het mogelijk om met meer zekerheid en dus stelligheid de eerdergenoemde ‘contribution claims’ en uiteindelijk het ‘contributieverhaal’ op te stellen. De combinatie van CA en PT is daarom te prefereren boven alléén CA of alléén PT.²²

3.2.3 Outcome mapping

Een derde aanpak die we hier aanstippen is outcome mapping. Deze aanpak wordt toegepast wanneer onderzoekers willen begrijpen welke uitkomsten zich hebben voorgedaan, zonder dat er op voorhand een gedetailleerde Theory of Change wordt opgesteld die de empirische analyses structureert. Meestal zijn de inputs bekend, en is er wellicht een idee over de uiteindelijk gewenste impacts, maar wordt er vervolgens een open houding gehanteerd in het analyseren van wat de inputs teweeg hebben gebracht – en via welke mechanismen dat verliep. Met de verkregen inzichten kan vervolgens een Theory of Change worden uitgewerkt.



Outcome mapping is dus vooral interessant als er op voorhand geen duidelijk beeld is van hoe ondernomen acties dienen uit te pakken en hoe ze daarbij afhankelijk zijn van elkaar. Als veranderpaden niet parallel aan elkaar lopen, maar op vele manieren met elkaar en met externe factoren verstrengeld zijn, dan is er sprake van complexe causaliteit.²⁷ Hoewel de effecten van meervoudige (interacterende) causale paden en condities lastig te voorspellen zijn, zijn ze wel te traceren. Bij evaluaties die dit pogen komt de nadruk te liggen op het in kaart brengen van ‘emergente’ resultaten. Hierbij wordt met een open blik gekeken welke verwachte en onverwachte ontwikkelingen zich voordoen nadat er beleidsinzet is gepleegd. Het is lastig om deze ontwikkelingen als beleidseffecten te bestempelen, doordat er geen solide methodologische basis is voor causale attributie. De aanpak geeft een beeld van wat er allemaal gebeurt, en hoe dat samenhangt, maar niet per se in hoeverre dat ook terug te herleiden is op de (beleids)interventie.

Deze aanpak lijkt voor de evaluatie van NGF-projecten in het algemeen wat minder relevant. Voor die projecten is vaak wel degelijk een Theory of Change beschikbaar. Hoewel die theorie mogelijk aangescherpt moet worden om up-to-date te zijn en nauwgezette CA en/of PT toe te laten, is dat startpunt voor doeltreffendheid- en doelmatigheidsevaluaties interessanter dan ongestructureerd veranderingen in kaart brengen. Bovendien zeggen de bevindingen van outcome mapping meer als een interventie al een geruime tijd loopt. Doordat er geen Theory of Change gebruikt wordt kun je in een vroeg stadium niet zeggen of het beleid op de goede weg zit, want de route die afgelegd zou moeten worden is niet bekend. Vermoedelijk duurt het bij NGF-projecten ook heel lang voordat kennisontwikkeling, onderzoek en innovatie resulteren in maatschappelijk relevante uitkomsten. Bij CA en PT is het wel mogelijk om tussentijds al te bepalen of er wenselijke voortgang is, namelijk door te kijken naar de eerste-orde uitkomsten die volgens de Theory of Change gaan doorwerken in effecten op hogere orde uitkomsten.

²⁷ Geels, F. W. (2022). Causality and explanation in socio-technical transitions research: Mobilising epistemological insights from the wider social sciences. *Research policy*, 51(6), 104537.

Hoofdstuk 4

Methode voor evaluatie NGF-projecten

In dit hoofdstuk wordt een voorstel gedaan voor een universele methode voor het evalueren van NGF-projecten.

4 Methode voor evaluatie NGF-projecten

In dit hoofdstuk doen we een voorstel voor een relatief universele methode voor het evalueren van de set van NGF-projecten. We beginnen met een logische selectie van evaluatieaanpakken die hiervoor in aanmerking komen (sectie 4.1). Dit resulteert in de keuze voor een methode die contributie-analyse en process tracing combineert. Vervolgens beschrijven we een protocol voor het toepassen hiervan (sectie 4.2). In sectie 4.3 reflecteren we op de geschiktheid van de voorgestelde methode.

Op basis van de in hoofdstuk 3 behandelde overwegingen en methodologische (on)mogelijkheden beschrijven we hieronder de contouren van een methode voor de evaluatie van NGF-projecten. Dit betreft een algemene opzet, en is dus enkel een voorstel op hoofdlijnen. Voor de verdere uitwerking is het nodig om de opzet concreet te maken voor de specifieke doelen, vormgeving en uitvoering van de afzonderlijke NGF-projecten. Een eerste verkenning hierop volgt in hoofdstuk 5 en 6, wat aanvullende aandachtspunten genereert die we in hoofdstuk 7 bespreken. Die aandachtspunten gaan niet alleen over de voorgestelde evaluatieaanpak en -methoden, maar ook over het proces dat voorafgaat aan het ontwerpen van een evaluatiestudie.

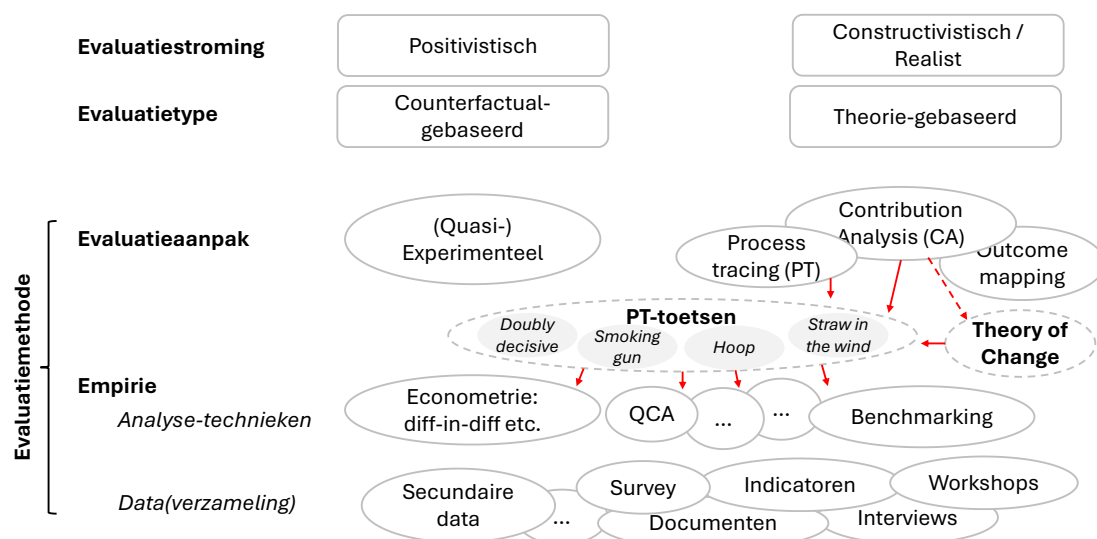
4.1 Logische selectie van passende evaluatieaanpakken

Welke opties er zijn bij het vormgeven van een evaluatiemethode hangt nauw samen met de kenmerken van de NGF-projecten. Die projecten bestaan vaak uit meerdere actielijnen, die worden ingezet om één of meerdere (eind)doelen na te streven. Verondersteld wordt dat die einddoelen een positieve maatschappelijke impact hebben. Sinds de derde investeringsronde, en voor enkele voorstellen uit de tweede ronde, is dit per project vastgelegd in een Theory of Change die onderdeel vormde van het projectvoorstel, en die in sommige gevallen al tussentijds bijgewerkt is. Actielijnen binnen een NGF-project kunnen dezelfde of verschillende doelgroepen hebben; causale paden kunnen met elkaar verstrengeld zijn; soms is er veel afstand tussen outputs en outcomes; etc. Hierdoor is er bij NGF-projecten sprake van complexe causaliteit; effect volgt uit de optelsom van in elkaar grijpende inspanningen en omstandigheden. Als er een duidelijke counterfactual is kun je van die complexiteit abstraheren, door te kijken of NGF-impulsen leiden tot extra inputs (input-additionaliteit) en extra outcomes (outcome-additionaliteit). Zoals besproken aan het begin van sectie 3.2 is er waarschijnlijk zelden een counterfactual op projectniveau.²⁸ Daarmee valt de mogelijkheid af om op quasi-experimentele wijze te onderzoeken of de NGF-impuls effect had op de uiteindelijk beoogde impact-maat, en hoe groot die was.

²⁸ Een uitzondering is wellicht het NGF-project 'Opschaling publiek private samenwerking in het beroepsonderwijs', waarbij een deel van de bestaande PPS'en in het NGF-project worden opgeschaald en de rest niet.

Wat wel nog kan is het 'openen van de black box' van werkingsmechanismen, door middel van outcome mapping (3.2.2) na te gaan welke ontwikkelingen zich hebben voorgedaan. Dit vergt echter dat de projecten al enkele jaren lopen, te meer omdat het behoorlijk wat tijd en stappen vergt voordat kennisontwikkeling, onderzoek en innovatie resulteren in maatschappelijke relevante uitkomsten. En zelfs al zijn we jaren verder: bij outcome mapping ligt de primaire focus niet op het analyseren van veronderstelde causale relaties op basis van een Theory of Change, maar op het identificeren van (veranderingen in) outcomes en het begrijpen hoe deze tot stand zijn gekomen. Voor een context waarin we [1] juist geïnteresseerd zijn in het bepalen van de doeltreffendheid en doelmatigheid van het NGF-project en de NGF-subsidie, en [2] er wél een Theory of Change beschikbaar is, lijkt outcome mapping niet de meest logische keuze te zijn.

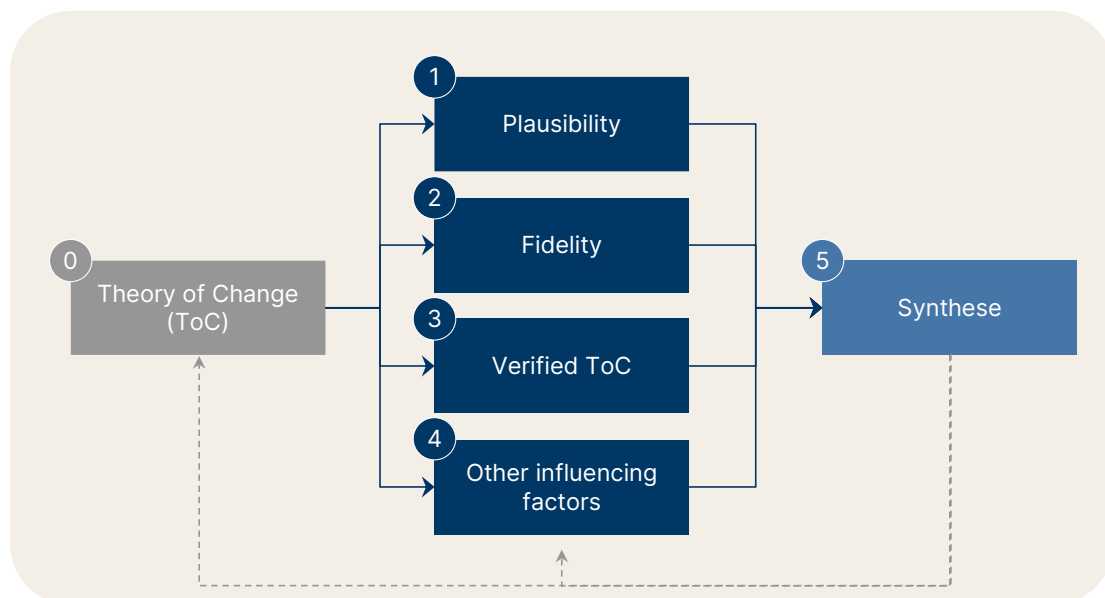
De optie die voorlopig overblijft is om gebruik te maken van contribution analysis en process tracing; twee aanpakken om op systematische wijze te toetsen of er empirisch bewijs is voor het zich voordoen van de in een Theory of Change beschreven causale mechanismen. De aanpakken zijn het sterkst als ze gecombineerd worden in één methode, waarbij process tracing de toetsen levert om zo betrouwbare causaliteitsclaims te kunnen doen. De gecombineerde aanpak is krachtig in een context waarin [1] er sprake is van parallelle actielijnen en veel (verschillende typen) betrokken partijen, [2] er niet gewerkt kan worden met duidelijke experimentele groepen en controlegroepen op het niveau van het gehele NGF-project, [3] beoogde uitkomsten niet altijd even goed te kwantificeren zijn, en [4] er een Theory of Change ligt die als basis kan dienen voor de evaluatie. Bij het empirisch toetsen van resultaten en schakels in de Theory of Change kan er gebruik worden gemaakt van verschillende soorten kwalitatieve en kwantitatieve data en evaluatietechnieken. Figuur 7 toont hoe de voorgestelde methode past in het in hoofdstuk 3 geïntroduceerde kader met evaluatiebegrippen.



Figuur 7: Visuele weergave van hoe de voorgestelde methode past in het kader met evaluatiebegrippen.

4.2 Stappen voor het toepassen van de voorgestelde methode

Als we een combinatie van contributie-analyse en process tracing willen aanwenden voor evaluatie van NGF-projecten, ligt het voor de hand om aan te sluiten bij de stappen en toetsen uit die respectievelijke aanpakken. Deze hebben we samengebracht in Figuur 8. De Theory of Change biedt de basis voor vier stappen die ontleend zijn aan de condities waar in contributieanalyse op getoetst moet worden. Daarbij is stap 4 (uitsluiten alternatieve verklaringen) nader uitgewerkt door de toetsen uit process-tracing te gebruiken voor het doen van zo betrouwbare uitspraken over causaliteit. De bevindingen uit de vier stappen samen vormen de input voor de synthese-stap. Die synthese resulteert in een gemotiveerd 'contributieverhaal', dat beschrijft welke bijdragen beleid heeft geleverd aan verwachte en onverwachte ontwikkelingen en hoe hard het bewijs daarvoor is.



Figuur 8: Stappen in evaluatiemethode die contributie-analyse en process tracing combineert.

Hieronder beschrijven we de stappen in iets meer detail.

0. Theory of Change opstellen of aanscherpen

In de context van de NGF-projecten gaan we ervan uit dat er in de voorstelfase al een eerste versie van de Theory of Change (ToC) is opgesteld, of dat daar (op korte termijn) aan wordt gewerkt. Binnen enkele NGF-projecten is men die ToC nu verder aan het bijwerken, bijvoorbeeld omdat het consortium in de voorstelfase nog algemeen is gebleven of omdat er op basis van feedback en voortschrijdend inzicht nieuwe ideeën zijn over de veranderepaden en externe factoren die van belang zijn. Aanscherpen van de ToC kan in ieder geval op de volgende manieren:

- Actualiseren: zijn er zaken gewijzigd sinds de voorstel-fase?

- Uitbreiden: zijn alle inputs, throughputs, outputs, outcomes, impacts en externe factoren (!) compleet?
- Concretiseren: wat behelzen de elementen in de Theory of Change precies? Des te specifieker dit gedefinieerd wordt, des te duidelijker het wordt welke indicatoren geschikt zijn om voortgang te monitoren.
- Opdelen: is het werkbaar om de Theory of Change als geheel te onderzoeken, of helpt het om deze op te delen in meer behapbare stukken? Bij grote uitgebreide programma's, zoals veel NGF-projecten, kan het helpen om een overkoepelende Theory of Change op te stellen, met voor specifieke onderdelen (bijvoorbeeld afzonderlijke werkpakketten) ook nog een meer gedetailleerde Theory of Change die in het overkoepelende raamwerk 'genest' is.

1. Toetsen op 'Plausibility' van de Theory of Change ('Aannemelijkheid')

Voor de meest actuele en uitgewerkte versie van de Theory of Change kan beoordeeld worden of deze aannemelijk is. Het gaat hier om de vraag of de opgestelde Theory of Change plausibel is: zijn de voorgestelde logische ketens ('impact pathways') geloofwaardig en goed onderbouwd?

Om aannemelijkheid van impact pathways te beoordelen kan een evaluator inspecteren of de Theory of Change wordt ondersteund met bronnen die causale verbanden aannemelijk maken. Dit kunnen studies zijn die bewijzen dat een bepaalde relatie bestaat. Evaluatoren zouden ook zelf naar bewijzen van die relaties op zoek kunnen gaan, hetgeen wel beduidend bewerkelijker is. In het geval van de NGF-projecten zijn er in de aanvraag- en toekenningsprocedure momenten waarop onafhankelijke experts zich al hebben uitgelaten over de onderbouwing van de voorstellen. Ook die bronnen kunnen benut worden bij het beoordelen van de aannemelijkheid van de Theory of Change.

2. Toetsen op 'Fidelity' van de Theory of Change ('Getrouwheid')

Zoals beschreven in sectie 3.2.1 verwijst het abstracte begrip 'Fidelity' naar de mate waarin werkelijk ondernomen activiteiten trouw zijn aan de Theory of Change. Het gaat dus om het nalopen van de inputs en activiteiten. Zijn die volgens plan ingezet en uitgevoerd, of is er op specifieke punten in (het linker deel van) de beleidstheorie sprake van meer/minder of andere inputs en activiteiten dan verwacht? Als geplande activiteiten een logische functie hadden in de beleidstheorie, maar toch niet zijn uitgevoerd, dan wordt het minder waarschijnlijk dat een beoogde uitkomst zich toch gaat voordoen.

In de context van de NGF-projecten ligt het voor de hand om in ieder geval de plannen en (voortgangs)verslagen van NGF-projecten te raadplegen. Voor duiding kan het helpen om ook te spreken met nauw betrokken stakeholders, zeker als het gaat om verklaringen voor herzieningen en tegenslagen of het identificeren van aanvullende

activiteiten (die mogelijk niet in officiële rapportages staan, maar wel van belang zijn voor de uiteindelijke doeltreffendheid).

3. Verificatie van de Theory of Change

De derde stap, en tevens toets, heeft betrekking op het nalopen (verificatie) van de Theory of Change. Leidend is de vraag of de keten van verwachte resultaten zich heeft gemanifesteerd en de causale aannames over deze keten hebben standgehouden. Deze analyse kan gezien worden als de kern van de methode. Het is deze stap waarin de expliciet beschreven (veronderstelde) causale relaties onderzocht worden.

Er zijn twee soorten onderdelen van de Theory of Change die hier van belang zijn:



Blokjes. De blokjes representeren de te onderzoeken concepten. We willen weten of er een causale relatie is tussen twee blokjes waar een pijltje tussen getekend is.

Pijltjes. Het pijltje, of 'schakel', representeert de (veronderstelde) causale relatie tussen de twee blokjes. De testen binnen process tracing richten zich op het onderzoeken of dit pijltje in de praktijk ook echt bestaat.

Om de Theory of Change te verifiëren is het allereerst noodzakelijk om informatie te hebben over de blokjes. Bestaat een blokje überhaupt (bijv. "is er sprake van kennisontwikkeling op een bepaald thema") en/of welke omvang heeft het (bijv. "hoeveel publicaties en promovendi zijn hiermee gemoeid?"). Dit is te onderzoeken door zoveel mogelijk gegevens te verzamelen over de elementen in de Theory of Change.

Voor de twee subsidieprojecten binnen het NGF zijn veel indicatoren en KPI's geformuleerd die overeenkomen met de Theory of Change, zie hoofdstukken 5 en 6. Binnen sommige departementale NGF-projecten is men momenteel bezig met het beter afstemmen van monitoringsinspanningen en -indicatoren op de eigen Theory of Change. De verwachting is echter dat evaluatoren bij een eerste evaluatie ook zelf op zoek moeten gaan naar gegevens die iets zeggen over elementen in de Theory of Change. Een voorbeeld van een type databron dat dan relevant kan zijn, zijn databases die ook iets zeggen over ontwikkelingen in andere domeinen, landen en perioden. Dit maakt het mogelijk om ontwikkelingen te benchmarken, om zo een beeld te vormen van wat er gebeurd is nadat activiteiten in het NGF-project in werking zijn getreden.

De beschikbare data per blokje dienen als *basis* voor het kunnen uitvoeren van process tracing, maar zijn op zichzelf niet voldoende om causale relaties (pijltjes) aan te tonen.

Om het bestaan van de pijltjes aan te tonen zijn toetsen nodig die bewijs kunnen leveren voor het bestaan van de veronderstelde causale relatie, en dus ook het uitsluiten van alternatieve verklaringen. Dit zijn de vier process tracing toetsen. Deze dienen te worden uitgevoerd voor iedere pijl/schakel in de Theory of Change.

Zoals beschreven in sectie 3.2.1 kan het voorkomen dat de toetsen vragen om het zoeken van aanvullende gegevens, bijvoorbeeld om de NGF-activiteiten en -resultaten tegen te benchmarken of om zelfs een controlegroep te identificeren. Daarbij, of daarnaast, kunnen er ook geavanceerde analysetechnieken worden benut. Zo is het goed voorstelbaar dat veel NGF-projecten investeren in kennisontwikkeling, wat vaak maar één van de vele actielijnen is en die er pas toe doet als (verderop in de impact pathway) de geproduceerde kennis ook echt toegepast wordt. Dat vergt wel dat die kennis überhaupt ontstaat. Observeren dat er meer kennis op het geprioriteerde onderwerp gepubliceerd wordt is dan al interessant, maar het wordt interessanter als dit kan worden afgezet tegen bijvoorbeeld ontwikkelingen op andere onderwerpen in hetzelfde vakgebied door de jaren heen. Door zo'n vergelijking te maken ontstaan mogelijkheden om, op dit specifieke onderdeel van de Theory of Change, deel-analyses te doen volgens (een benadering van) de 'doubly decisive'-toets.

4. Afwegen van bevindingen tegen andere beïnvloedende factoren

De vierde stap bestaat uit het identificeren van andere factoren die van invloed kunnen zijn op de in voorgaande stappen onderzochte veranderpaden.

Een deel hiervan zal al opgenomen zijn in de Theory of Change. Op het moment dat hier ook activiteiten op gericht worden zijn de 'andere factoren' niet langer meer extern. Denk aan (supra)nationaal beleid dat van belang kan zijn voor het succes van een NGF-project. Wanneer dit erkend wordt, en leidt tot activiteiten om dat te beïnvloeden, wordt het (supra)nationale beleid zelf ook een element in de Theory of Change (deze moet immers veranderen om de gewenste impact te verhogen).

Er zijn ook andere factoren waarop geen actie ondernomen zal worden, maar die wel van belang zijn. Hiervoor dient te worden bepaald hoe groot dat belang is, en wat dus nog de relatieve invloed is van de acties die wel door het (NGF-)beleid beïnvloed worden. Soms kan dit verkend worden door te kijken of de omvang van verschillende invloeden te schatten en te vergelijken is, bijvoorbeeld als het gaat om financiële middelen (e.g. hoe verhoudt de NGF-investering in productiecapaciteit zich tot exogene investeringen vanuit andere partijen?). In andere gevallen is zo'n vergelijking niet mogelijk, en dienen evaluatoren zich vooral op interviews te baseren bij het identificeren en wegen van relevante andere factoren. Daarbij is het verstandig om te kijken naar de mate waarin onderbouwing en schattingen van verschillende gesprekspartners overeenkomen, aangezien dit een indicatie is dat ze betrouwbaarder zijn dan wanneer iedereen een heel ander beeld heeft van andere beïnvloedende factoren.

5. Synthese

Na het doorlopen van de stappen 1 t/m 4 zoals hiervoor beschreven is de cyclus van de Contribution Analysis doorlopen en kunnen de bevindingen samengebracht worden in deze vijfde en laatste synthesestap. In deze stap wordt een integraal verhaal van de contribution analysis neergezet, waarbij aangegeven kan worden welke impact pathways zich wel/niet voltrokken hebben en waar mogelijk een duiding hiervan. Dit verhaal wordt ook wel de contribution story genoemd; een krachtige samenvatting van het beeld dat ontstaat uit de gecombineerde bevindingen voor alle onderzochte elementen van de ToC. Bij contributieanalyse stellen evaluatoren eerst een ruwe versie op van de contribution story. Op basis van kritieken van anderen en van henzelf scherpen ze die vervolgens op iteratieve wijze verder aan, waarbij ze voor slecht onderbouwde onderdelen eventueel ook aanvullend onderzoek moeten doen.

In deze synthesestap worden alle bevindingen systematisch samengevat, door te laten zien waar bewijs is gevonden voor causale relaties in de ToC. Een overzichtsstudie van CA's in de praktijk suggereert om een 'Evidence Analysis Table' op te stellen, waarbij per causale link wordt genoteerd wat voor data is gebruikt, wat de bevinding is t.a.v. de verwachte relatie (aangetoond of niet aangetoond), wat de aard is van het causale mechanisme waar de relatie betrekking op heeft (bijv. een beoogde bijdrage, of slechts een conditie die van belang is voor een beoogde bijdrage), en hoe sterk het empirische bewijs is.²⁹

Naast tabelvorm zijn er ook andere manieren om de synthese visueel vorm te geven. In hoofdstukken 5 en 6 geven we voor de twee concrete NGF-projecten voorbeelden voor een schets van een fictieve uitkomst waarin getoond wordt welke onderdelen (blokjes) wel/niet bereikt zijn en voor welke causale mechanismen (pijltjes) wel/geen bewijs is gevonden. Op basis van de contribution analysis kan dus aangegeven worden welke impact pathways wel/niet geslaagd zijn inclusief een toelichting op basis van de verzamelde informatie.

Deze aanpak biedt dus ook de ruimte om een genuanceerd beeld van de doeltreffendheid neer te zetten. Het hoeft niet zo te zijn dat het gehele NGF-project wel of niet doeltreffend is geweest. Het kan zijn dat het in termen van bepaalde impact pathways wel doeltreffend is geweest en in termen van andere impact pathways niet. Daar komt ook bij dat er externe factoren zijn die [1] bepalend zijn en [2] niet of nauwelijks te beïnvloeden zijn geweest. Het kan in theorie zo zijn dat het NGF-project slim is ingericht *gegeven de informatie die men bij de start had*, maar dat het later in de tijd toch niet tot de gewenste resultaten heeft geleid. Het feit dat het project niet doeltreffend zou blijken is niet hetzelfde als zeggen dat het niet slim is geweest om de kans bij de start

²⁹ Delahais, T. and Toulemonde, J. (2012) 'Applying Contribution Analysis: Lessons from Five Years of Practice', *Evaluation* 18.3: 281-93,

te wagen. Dergelijke nuances zijn middels deze aanpak goed aan te brengen, omdat de Theory of Change een basis biedt om deze argumenten gestructureerd in te bedden in de 'logica van het project'.

4.3 Geschiktheid van voorgestelde methode

Het toepassen van theorie-gebaseerde evaluatieaanpakken, zoals contribution analysis, al dan niet in combinatie met process tracing, is aantrekkelijk als het gaat om haalbaarheid: er is immers een heel scala van meer en minder bewerkelijke manieren om een *Theory of Change* na te lopen. De kwaliteit van de analyse hangt af van keuzen inzake mate van detail van de ToCs; de diversiteit aan databronnen; en de scherpheid van de tests. Het advies is uiteraard om een zo hoog mogelijk mate van causale betrouwbaarheid na te streven. Binnen CA is het mogelijk om econometrie toe te passen om te toetsen of één of meerdere impactpaden zich ook als zodanig manifesteren. Econometrische methoden zijn ook te zien als een (zeer) strenge toets om causaliteit te onderbouwen op een bepaalde veronderstelde *impact pathway*.

Voorwaarde (of eerste stap) voor CA/PT-analyse is het in kaart brengen van de Theory of Change. De exercitie met het opstellen, aanscherpen en/of uitbreiden van de Theory of Change is **waardevol voor de departementen en consortia zelf**. Vaak willen ze weten of ze de juiste dingen doen (onder omstandigheden die zich blijven ontwikkelen) en of ze goed op weg zijn. Zo'n oefening biedt ook een basis voor het **vaststellen van relevante indicatoren**. Dit vergroot de kans op medewerking (belasting voor Theory of Change aanscherpen is strategisch i.p.v. administratief, en de werkelijke administratieve lasten zijn later meer de moeite).

Tenslotte is het relevant om te benadrukken dat de voorgestelde aanpak geschikt is om op zoek te gaan naar zo aannemelijk mogelijke conclusies over de causale invloed van interventies (attributie), én om ondertussen ook te leren waar(om) effecten wel/niet optreden.

Hoofdstuk 5

Evaluatieontwerp Groeien met Groen Staal

In dit hoofdstuk wordt een mogelijk evaluatieontwerp voor NGF-subsidieproject 'Groeien met Groen Staal' gepresenteerd.

5 Evaluatieontwerp Groeien met Groen Staal

5.1 Inleiding

In dit hoofdstuk presenteren we het evaluatieontwerp voor het bepalen van de doeltreffendheid en doelmatigheid van het NGF-project 'Groeien met Groen Staal' (GGS). De eerste evaluatie staat gepland voor 2026. Aangezien het project loopt tot en met 2031, betekent dit dat het project nog niet volledig is uitgevoerd en niet alle beoogde uitkomsten al een kans hebben gehad om zich te manifesteren. Concreet kan in 2026 dus al wel wat gezegd worden over de voorlopige resultaten, maar kan nog geen sluitende evaluatie uitvoeren ten aanzien van de doeltreffendheid en doelmatigheid van het gehele NGF-project.

In dit hoofdstuk wordt een evaluatieontwerp voor het gehele NGF-project gepresenteerd. Het hoofdstuk wordt afgesloten met een reflectie op wat naar verwachting reeds mogelijk zal zijn in 2026, en wat mogelijk zal zijn in 2031 en daarna (0).

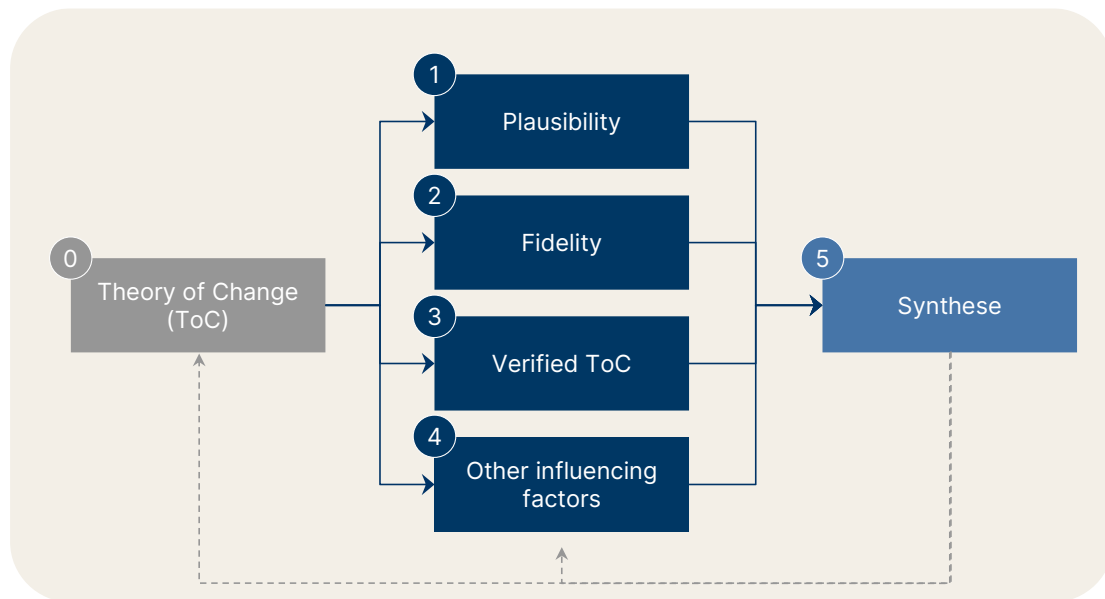
Het evaluatieontwerp wordt opgebouwd aan de hand van de volgende bouwstenen:

Tabel 5: Opzet evaluatieontwerp GGS

Onderdeel evaluatieontwerp	Sectie
Doeltreffendheid GGS	5.2
0. Theory of Change	5.2.1
1. Plausibility	5.2.2
2. Fidelity	5.2.3
3. Verified Theory of Change	5.2.4
4. Other influencing factors	5.2.5
5. Synthese stappen 1 t/m 4	5.2.6
Doelmatigheid	5.3
Vooruitblik	0

5.2 Doeltreffendheid GGS

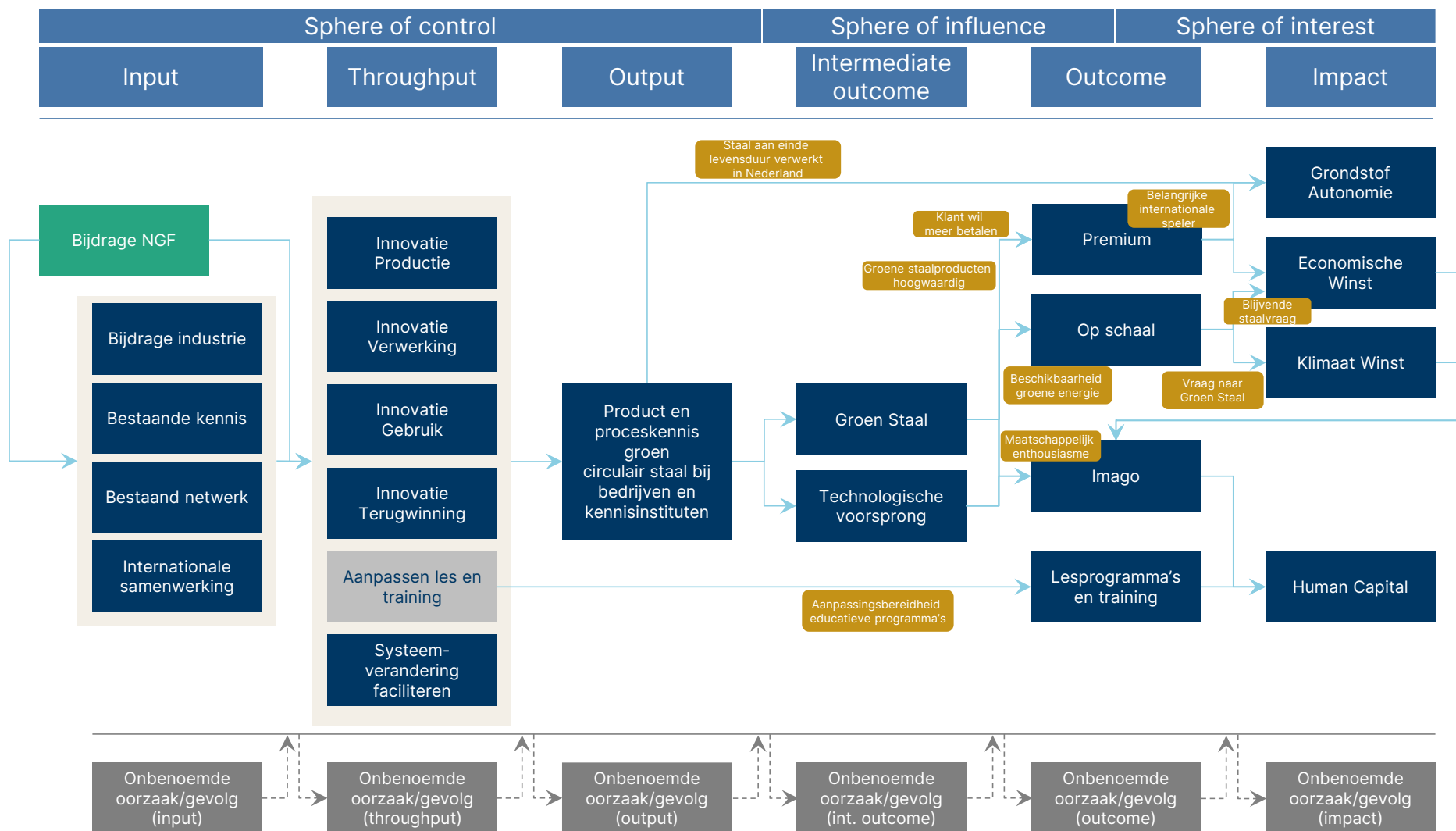
Om de doeltreffendheid van GGS (en de inzet van de NGF-middelen) te beoordelen volgen we de stappen uit de in hoofdstuk 4 beschreven methode:



Figuur 9: Stappen in evaluatiemethode die contributie-analyse en process tracing combineert.

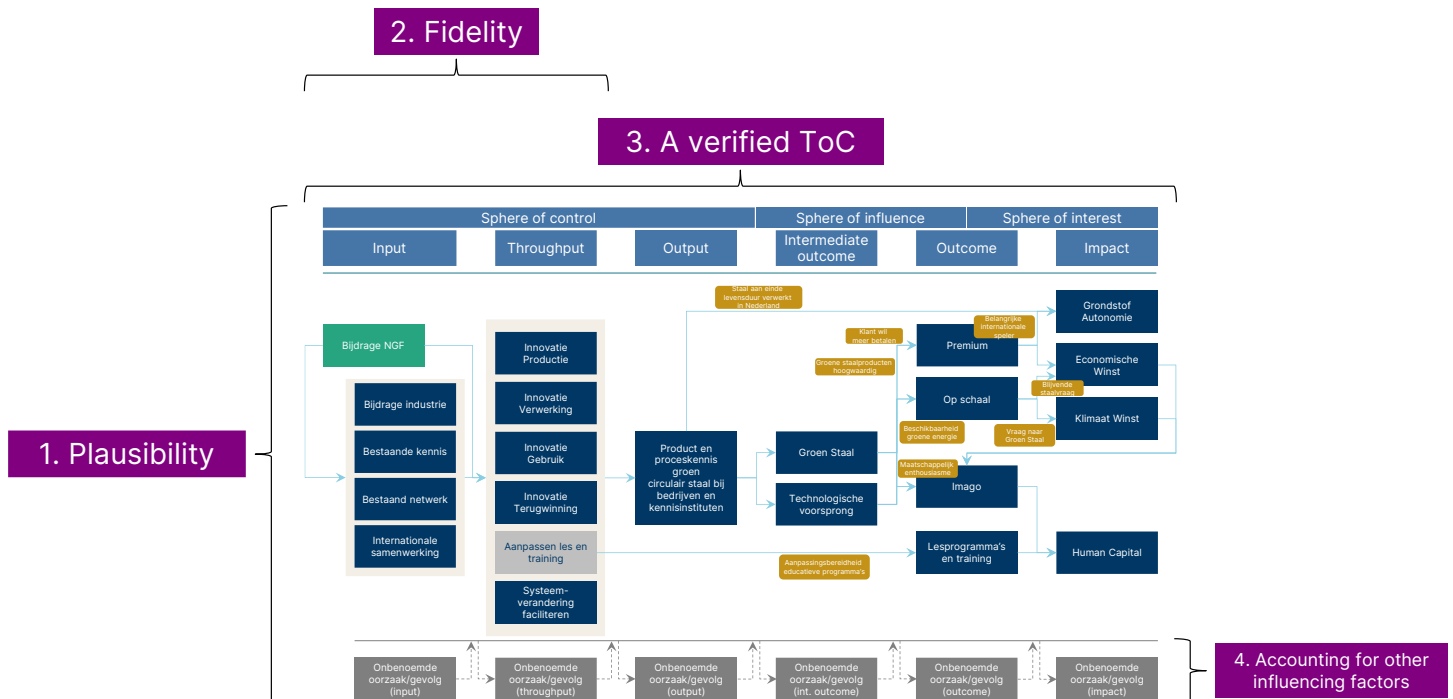
5.2.1 [0] Theory of Change (ToC)

De Theory of Change zoals deze in het projectplan van de subsidieontvangers is gepresenteerd vormt de basis voor de contribution analysis. In Figuur 10 wordt deze Theory of Change weergegeven. Voor meer informatie over de (inhoudelijke) opzet van dit NGF-project verwijzen we terug naar sectie 2.2.



Figuur 10: Theory of Change GGS. Bron: projectplan GGS, bewerking Dialogic

De vier condities waar binnen een contribution analysis aan voldaan moet worden om causaliteit vast te stellen, kunnen als volgt visueel aan de ToC gerelateerd worden:



Figuur 11 De vier criteria geplot op de Theory of Change

5.2.2 [1] Plausibility

Het eerste criterium om naar te kijken is 'Plausibility'.

Voor de evaluatie stellen wij voor om in beginsel te **vertrekken vanuit de eerdere expertoordelen** ten aanzien van het NGF-project. Het beoordelen van de plausibiliteit van de Theory of Change vergt immers substantiële expertise op het gebied van de staalindustrie, innovatie en beleid. Hoewel de toekomstige evaluatoren hier potentieel zelf ook een kritische reflectie op kunnen bieden, borgt het aansluiten bij bestaande expertoordelen dat er een solide basis is om de evaluatie goed uit te kunnen voeren.

De centrale bron is de **beoordeling van de Adviescommissie Nationaal Groeifonds**³⁰. De commissie heeft alle inputs van (externe) experts meegenomen en is tot een integraal oordeel gekomen dat in deze bron wordt beschreven. Ten aanzien van de strategische onderbouwing wordt gesteld *"De commissie vindt de strategische onderbouwing van het project aannemelijk"* en over de kwaliteit wordt gezegd *"De kwaliteit van het voorstel wordt door de commissie als voldoende beoordeeld"*. Een specifiek punt waar de commissie een bedenking bij heeft is het onderdeel 'Terugwinning',

³⁰ Commissie Nationaal Groeifonds (juni 2023), Rapport derde beoordelingsronde

waarbij zij aangeeft “Voor het onderdeel Terugwinning lijkt de voorgestelde opschaalingsstap te groot en de financiële onderbouwing daarvan te mager.”.

Een belangrijke bron van input voor de commissie is het **oordeel van het CPB**. Het CPB heeft een analyse uitgevoerd op het project en heeft de resultaten gepubliceerd³¹. De commissie heeft gereageerd op de aandachtspunten van het CPB. Het CPB is op bepaalde punten minder positief dan de commissie, met name waar het aankomt op de legitimiteit van de publieke investeringen en de mate waarin het reëel is dat de beoogde outcomes en impact toe te schrijven zijn aan het voorstel.

b. Is aannemelijk dat de initiatieven de problemen verhelpen (gegeven de bij ons beschikbare kennis over het ecosysteem)?	Overall beeld Dit is deels niet aannemelijk, deels niet te beoordelen. We motiveren dit met twee algemene en drie specifieke bevindingen. • Algemeen risico: Het voorstel geeft aan dat er gedurende de periode voorgestelde demonstrators (en dus onderdelen van het voorstel) gestaakt kunnen worden door verandering in onder meer de bedrijfsstrategie. • Algemene kanttekening: Het voorstel formuleert concrete en ambitieuze doelen, maar deze zijn of niet gekoppeld aan maatregelen, of niet aan maatregelen die binnen het voorstel vallen. Specifieke bevindingen: • Via de innovaties binnen thema II (productie) van het voorstel kan het voorstel in potentie een bijdrage leveren aan deelprobleem 1 (klimaat) in de vorm van kennisontwikkeling • Of thema IV (gebruik) bijdraagt aan deelprobleem 3 (inefficiënt hergebruik van staal), is niet te beoordelen door een gebrek aan uitwerking van de onderdelen. • Er bestaat al technologie om luchtvervuiling en CO₂-uitstoot bij staalproductie substantieel terug te dringen, wat additionele kennisbaten van nieuw onderzoek hiernaar substantieel beperkt. Het voorstel beschouwt deze bestaande technologieën echter niet. Of en hoeveel de technologieën in het voorstel luchtvervuiling en CO₂-uitstoot zouden kunnen terugdringen, wordt bovendien niet duidelijk.
--	---

Figuur 12 Koppeling activiteiten aan doelen. Bron: CPB-analyse voorstellen Nationaal Groeifonds

Uitgaande van het oordeel van de commissie kan op dit moment gesteld worden dat de onderbouwing van het plan en daarmee de Theory of Change over het geheel gezien plausibel is. Er zijn weliswaar bedenkingen over hoe realistisch bepaalde ‘impact pathways’ zijn, maar een evaluatie kan hier juist expliciet aandacht voor hebben. In de beoogde evaluatie kunnen bovenstaande reflecties aangevuld worden met aanvullende (nieuwe) reflecties en bespiegelingen om een oordeel te vellen over de plausibiliteit van de Theory of Change.

Daarbij kunnen de evaluatoren de 17 criteria voor het toetsen van de kwaliteit van een Theory of Change meenemen in hun oordeel over de plausibiliteit, zie 4.2.

³¹ CPB-analyse voorstellen Nationaal Groeifonds (2023), Derde beoordelingsronde, 2023

5.2.3 [2] Fidelity

Het tweede criterium gaat over de '**Fidelity**', ofwel: zijn de acties zoals beschreven in de Theory of Change ook daadwerkelijk als zodanig uitgevoerd?

We baseren ons hier in beginsel op de subsidiebeschikking, het projectplan en de aankomende **voortgangsrapportages** waarin de voortgang op de activiteiten (per werkpakket) worden beschreven. Op basis van de voortgangsrapportages kan vastgesteld worden of de acties inderdaad zijn uitgevoerd zoals beschreven.

In dit NGF-project is reeds bekend dat er wijzigingen plaatsvinden op het onderdeel 'Innovatie Terugwinning', omdat de partij die aan de slag zou gaan met scheepsrecycling niet aan de voorwaarden kon voldoen. Voor de evaluatie in 2026 kunnen de evaluatoren er alvast rekening mee houden dat er op dit punt een wijziging is ten opzichte van het oorspronkelijke plan.

In generieke zin geldt dat de evaluatoren oog zullen moeten hebben voor wijzigingen ten opzichte van het oorspronkelijke projectplan. Deze wijzigingen kunnen op hun beurt ook doorwerken in de rest van de Theory of Change. Het is daarom nuttig om **voorafgaand aan de evaluatie in beeld te hebben wat de wijzigingen zijn en wat de verwachte doorwerkingseffecten hiervan zijn in de Theory of Change**. Als er al duidelijke verwachtingen zijn waarom bepaalde genoemde effecten niet/minder plaats zullen vinden en/of bepaalde nieuwe effecten kunnen gaan optreden, kan de Theory of Change die als basis dient voor het evaluatieontwerp aangepast worden.

5.2.4 [3] Verified Theory of Change

Het derde criterium heeft betrekking op een '**Verified Theory of Change**', ofwel: kunnen we bewijs vinden voor de veronderstelde causale relaties in de Theory of Change?

Voor ieder blokje in de Theory of Change is geïnventariseerd welke data en indicatoren beschikbaar zijn of worden. Het volledige overzicht hiervan is als bijlage beschikbaar voor EZ en RVO, maar kan omwille van vertrouwelijkheid niet gepubliceerd worden. In de tabel hieronder is een overzicht gegeven van het aantal indicatoren per onderdeel in de ToC.

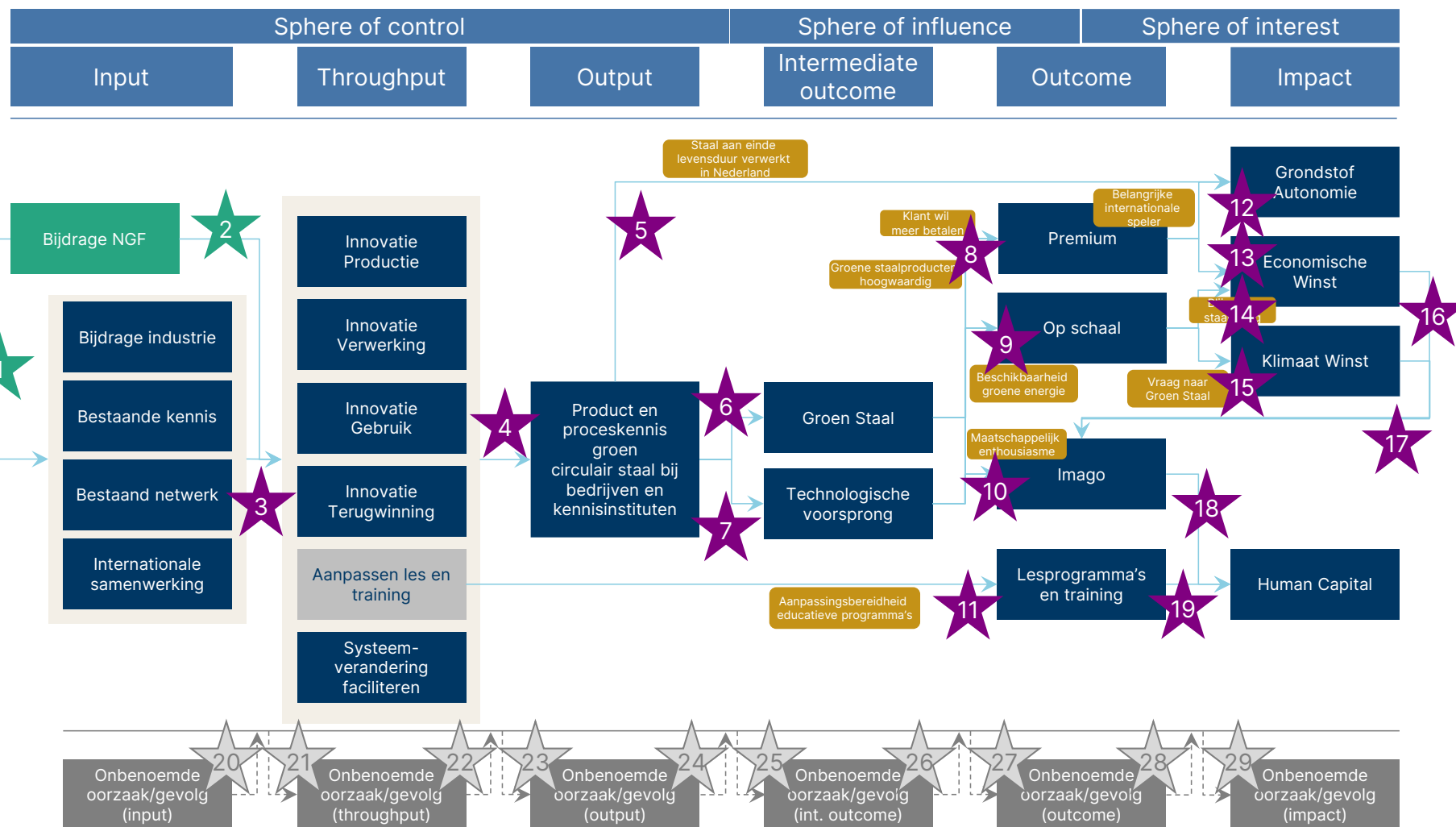
Tabel 6: Aantal beschikbare indicatoren per onderdeel in de ToC

Onderdeel van de ToC	Aantal indicatoren
Input	5
Throughput	28
Output	83
Int. Outcome	8

Onderdeel van de ToC	Aantal indicatoren
Outcome	19
Impact	4
Totaal	147

Het is een groot aantal indicatoren, waarbij iedere indicator gerichte informatie geeft over een specifiek onderdeel van het projectplan. Voor het evalueren van de doeltreffendheid is dit wenselijk, omdat hiervoor de causale relaties in het plan zo concreet en gedetailleerd mogelijk onderzocht dienen te worden. Omwille van de (tussentijdse) evaluatierapportage kan ervoor gekozen om (ook) een aantal geaggregeerde indicatoren te presenteren, zoals 'het percentage van de milestones dat behaald is'. Een dergelijke geaggregeerde indicator kan in één oogopslag wat zeggen over hoe het staat met de voortgang, maar kan tegelijkertijd dus niet aangeven wélke voortgang exact geboekt is.

Om op basis van die indicatoren iets te zeggen over causale relaties, is het zaak om ook de pijltjes (schakels) in de Theory of Change te onderzoeken en te toetsen. In Figuur 13 worden de pijltjes genummerd (met sterretjes) waarvoor een process tracing test opgesteld is.



Figuur 13: De 'impact pathways' die getest worden in de evaluatie van GGS

Zoals in sectie 3.2 is beschreven zijn er vier typen testen, te weten Straw-in-the-Wind, Hoop, Smoking Gun, en Doubly decisive. Hieronder worden de voorgestelde testen voor GGS beschreven. De nummering is conform de nummering in de sterren in Figuur 13. De beschrijvingen van de testen zijn omwille van vertrouwelijkheid gepseudonimiseerd conform de nummering van de indicatoren die overeenkomen met de bijlage die beschikbaar is voor EZ en RVO.

Tabel 7 Process tracing tests voor GGS

ID ToC	Type test	Type criterium	Beschrijving test	Uitvoering test (dataverzamelmethode en data-bron)
1	Hoop	Noodzakelijke voorwaarde	Hebben betrokken partijen vanaf de start van het NGF-traject meer tijd en middelen aangewend om te investeren in Groen Staal? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, dan is de hypothese dat de NGF-bijdrage leidt tot meer input verworpen	- Deskresearch: voortgangsrapportage RVO, projectplan, expertbeoordeling projectplan, jaarverslagen deelnemers project - Interviews: bevragen betrokken stakeholders en (externe) experts over input-additionaliteit
2	Hoop	Noodzakelijke voorwaarde	Worden er binnen GGS activiteiten ondernomen die zonder NGF-bijdrage niet/minder/anders/later ondernomen zouden zijn? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, dan is de hypothese dat de NGF-bijdrage leidt tot meer/andere throughput verworpen	- Deskresearch: voortgangsrapportage RVO, projectplan, expertbeoordeling projectplan, jaarverslagen deelnemers project - Interviews: bevragen betrokken stakeholders en (externe) experts over input-additionaliteit
3	Doubly Decisive	Doorslaggevend bewijs	Worden de voorgestelde activiteiten uitgevoerd op basis van de benoemde inputs? - Zo ja, bevestigt hypothese en verworpt hypothese dat activiteiten (enkel) door andere inputs tot stand zijn gekomen - Zo nee, verworpt hypothese dat activiteiten tot stand zijn gekomen (mede) vanwege de benoemde inputs	- Deskresearch: voortgangsrapportage RVO - Interviews: bevragen betrokken stakeholders
4	Doubly Decisive	Doorslaggevend bewijs	Worden de voorgestelde outputs (milestones en deliverables) gecreëerd op basis van de benoemde activiteiten? - Zo ja, bevestigt hypothese en verworpt hypothese dat outputs (enkel) door andere inputs tot stand zijn gekomen - Zo nee, verworpt hypothese dat outputs tot stand zijn gekomen (mede) vanwege de benoemde activiteiten	- Deskresearch: voortgangsrapportage RVO - Interviews: bevragen betrokken stakeholders
5	Hoop	Noodzakelijke voorwaarde	Wordt er in de demonstrators Groen Staal gemaakt waarbij minder benodigde grondstoffen worden geïmporteerd dan bij de productie van niet-groen/'traditioneel staal? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast	- Deskresearch: voortgangsrapportage RVO, documentatie bij stakeholders over productie (groen) staal, jaarverslagen staalproducenten - Interviews: bevragen betrokken stakeholders over productie

			- Zo nee, dan is de hypothese dat de output leidt tot (meer) grondstofautonomie verworpen	
6	Doubly Decisive	Doorslaggend bewijs	Wordt er [1] Groen Staal gemaakt c.q. indicatoren #117 tot en met #123 worden behaald waarbij [2] kennis en/of resultaten worden gebruikt die enkel binnen/vanwege het NGF-project tot stand zijn gekomen? - Zo ja, bevestigt hypothese en verworpt hypothese dat Groen Staal (enkel) door andere oorzaken tot stand is gekomen. - Zo nee, verworpt hypothese dat Groen Staal tot stand is gekomen vanwege het NGF-project	- Deskresearch: voortgangsrapportage RVO, documentatie bij stakeholders over productie (groen) staal - Interviews: bevragen betrokken stakeholders over productie
6	Hoop	Noodzakelijke voorwaarde	Wordt er Groen Staal gemaakt c.q. indicatoren #117 tot en met #123 worden behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, dan is de hypothese dat het NGF-project leidt tot Groen Staal verworpen	- Deskresearch: voortgangsrapportage RVO, documentatie bij stakeholders over productie (groen) staal - Interviews: bevragen betrokken stakeholders over productie
7	Hoop	Noodzakelijke voorwaarde	Heeft de Nederlandse staalindustrie op het gebied van Groen Staal een technologische voorsprong op andere landen zoals Zweden, Oostenrijk, Duitsland, China, India, Japan, VS en Rusland? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, dan is de hypothese dat het NGF-project leidt tot een technologische voorsprong verworpen	Benchmarking NL met andere landen o.b.v.: - Deskresearch: voortgangsrapportage RVO, marktbeschrijvingen, jaarrapportages grote staalproducenten, ... - Interviews: bevragen betrokken stakeholders
8	Hoop	Noodzakelijke voorwaarde	Kan er gerekend worden op premium "groene" staalproducten c.q. indicatoren #125 tot en met #127 worden behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verworpt hypothese dat Groen Staal tot hogere marges leidt	- Deskresearch: eindrapportage RVO, jaarrekeningen staalproducenten, marktanalyses - Interviews: staalproducenten en afnemers
9	Doubly Decisive	Doorslaggend bewijs	Worden [1] de KPI's m.b.t. 'Op schaal' behaald conform indicatoren #128 tot en met #136 en [2] wordt hiervoor 'Groen Staal' zoals in het NGF-project ontwikkeld gebruikt? - Zo ja, bevestigt hypothese en verworpt hypothese dat 'Op schaal' (enkel) door andere oorzaken tot stand is gekomen. - Zo nee, verworpt hypothese dat 'Op schaal' tot stand is gekomen vanwege Groen Staal	- Deskresearch: eindrapportage RVO, jaarrekeningen staalproducenten, marktanalyses - Interviews: betrokken stakeholders

10	Hoop	Noodzakelijke voorwaarde	Is het imago van de staalindustrie in Nederland beter dan haar imago in 2024 gemeten door indicatoren #137 tot en met #139? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, dan is de hypothese dat Groen Staal leidt tot een beter imago van de staalindustrie verworpen	- Deskresearch: eindrapportage RVO, instroomcijfers DUO, werknemerstevredenheidsonderzoek, marktanalyses - Interviews: betrokken stakeholders
11	Hoop	Noodzakelijke voorwaarde	Zijn de vier KPI's met betrekking tot de lesprogramma's en training behaald c.q. zijn indicatoren #140 tot en met #143 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, dan is de hypothese dat het NGF-project leidt tot de beoogde lesprogramma's en training verworpen	- Deskresearch: eindrapportage RVO, opleidingsinformatie DUO, opleidingsgidsen mbo - Interviews: betrokken stakeholders
12	Hoop	Noodzakelijke voorwaarde	Is er sprake van grondstofautonomie door de Nederlandse staalindustrie c.q. is indicator #144 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, dan is de hypothese dat de premium leidt tot meer grondstofautonomie verworpen	- Deskresearch: eindrapportage RVO, jaarrekeningen staalproducenten, marktanalyses - Interviews: betrokken stakeholders
13	Hoop	Noodzakelijke voorwaarde	Is er sprake van een economische winst door de Nederlandse groene staalindustrie c.q. wordt indicator #145 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, dan is de hypothese dat de premium leidt tot economische winst verworpen	- Deskresearch: eindrapportage RVO, jaarrekeningen staalproducenten, marktanalyses, CBS Statline - Interviews: betrokken stakeholders - CBS-microdata: lijst van betrokken bedrijven + ABR + Productiestatistieken + BTW + NFO
14	Hoop	Noodzakelijke voorwaarde	Is er sprake van een economische winst door de Nederlandse groene staalindustrie c.q. wordt indicator #145 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, dan is de hypothese dat de premium leidt tot economische winst verworpen	- Deskresearch: eindrapportage RVO, jaarrekeningen staalproducenten, marktanalyses, CBS Statline - Interviews: betrokken stakeholders - CBS-microdata: lijst van betrokken bedrijven + ABR + Productiestatistieken + BTW + NFO
15	Doubly Decisive	Doorslaggevend bewijs	Is er sprake van een klimaatwinst door het gebruik van Groen Staal op schaal c.q. wordt indicator #146 behaald? - Zo ja, bevestigt hypothese en verwerpt hypothese dat 'Klimaatwinst' (enkel) door andere oorzaken tot stand is gekomen. - Zo nee, verwerpt hypothese dat 'Klimaatwinst' tot stand is gekomen vanwege het gebruik van Groen Staal op schaal.	- Deskresearch: eindrapportage RVO, jaarrekeningen staalproducenten, marktanalyses - Interviews: betrokken stakeholders

16	Hoop	Noodzakelijke voorwaarde	Is het imago van de staalindustrie in Nederland beter dan haar imago in 2024 gemeten door indicatoren #137 tot en met #139? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, dan is de hypothese dat Groen Staal leidt tot een beter imago van de staalindustrie verworpen	- Deskresearch: eindrapportage RVO, instroomcijfers DUO, werknemerstevredenheidsonderzoek, marktanalyses - Interviews: betrokken stakeholders
17	Hoop	Noodzakelijke voorwaarde	Is het imago van de staalindustrie in Nederland beter dan haar imago in 2024 gemeten door indicatoren #137 tot en met #139? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, dan is de hypothese dat Groen Staal leidt tot een beter imago van de staalindustrie verworpen	- Deskresearch: eindrapportage RVO, instroomcijfers DUO, werknemerstevredenheidsonderzoek, marktanalyses - Interviews: betrokken stakeholders
18	Hoop	Noodzakelijke voorwaarde	Is er sprake van een verbetering in Human Capital c.q. wordt indicator 147 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, dan is de hypothese dat het (verbeterde) imago leidt tot stabiele instroom van talent verworpen	- Deskresearch: eindrapportage RVO - Interviews: betrokken stakeholders - CBS-microdata: lijst van betrokken bedrijven + ABR + SPOLISBUS (+ DIPLOMAMBOTAB + DIPLOMAHOTAB)
19	Doubly Decisive	Doorslaggevend bewijs	Is er sprake van een verbetering in Human Capital c.q. wordt indicator 147 behaald? - Zo ja, bevestigt hypothese en verworpt hypothese dat 'Human Capital' (enkel) door andere oorzaken tot stand is gekomen. - Zo nee, verworpt hypothese dat 'Human Capital' tot stand is gekomen vanwege de lesprogramma's en training	- Deskresearch: eindrapportage RVO - Interviews: betrokken stakeholders - CBS-microdata: lijst van betrokken bedrijven + lijst van CREBO- en CROHO-codes van ontwikkelde/aangepaste opleidingen + ABR + SPOLISBUS + DIPLOMAMBOTAB + DIPLOMAHOTAB

Zoals de tabel laat zien kunnen de testen variëren in de manier waarop er invulling aan gegeven wordt in termen van de bijbehorende analysetechniek, dataverzamelingstechniek en databronnen. Enkele voorbeelden ter illustratie zijn:

- Voor de test bij **ster #6** wordt bepaald of Groen Staal (mede) door het NGF-project tot stand is gekomen. Hier kan in detail gekeken worden welke outputs er (enkel) binnen het project zijn opgeleverd en hoe die worden gebruikt in Groen Staal. Een combinatie van het bestuderen van de **voortgangsrapportages, documentatie bij stakeholders** over productie (groen) staal en **interviews** met betrokkenen kan hier licht op werpen.
- Voor de test bij **ster #7** kan gebruikgemaakt worden van **benchmarking** om te bepalen of de Nederlandse staalindustrie op het gebied van Groen Staal een technologische voorsprong heeft op andere landen zoals Zweden, Oostenrijk, Duitsland, China, India, Japan, VS en Rusland. Hiervoor kunnen bronnen binnen

het project (bijv. voortgangsrapportage) en bronnen buiten het project (bijv. marktrapportages) ingezet worden.

- Voor de test bij **ster #19** kan binnen de **microdata-omgeving van het CBS** onderzocht worden of er vanuit de nieuwe/aangepaste opleidingen instroom is in de sector. Hiervoor kunnen lijsten van opleidingscodes (CREBO en CROHO) en relevante organisaties (bijv. o.b.v. KVK-nummer) gecombineerd worden met bestaande registraties binnen het CBS, zoals het Algemene Bedrijvenregister en de SPOLISBUS waarin informatie over banen en lonen is opgenomen.

In de tabel is tevens te lezen dat **soms dezelfde test voor verschillende impact pathways** (pijltjes) gebruikt wordt. Dat is bijvoorbeeld het geval bij de pijltjes die naar 'Imago' leiden. Op dit punt testen we met dit ontwerp enkel op de noodzakelijke voorwaarde, omdat het lastig is om vast te stellen hoe het integrale concept van 'imago' exact beïnvloed is door de individuele blokjes 'Groen Staal, Technologische voorsprong', 'Economische winst' en 'Klimaatwinst'. Met diepgravend onderzoek zou dat mogelijk onderzocht kunnen worden, maar dat voert vermoedelijk te ver voor de beoogde evaluatieonderzoeken.

Voor veel impact pathways is een Hoop-test geformuleerd. Met een dergelijke test kan in feite de hypothese dat er wel een causale relatie is enkel gefalsificeerd worden. Simpel gezegd: als we uitkomsten waarnemen zoals beschreven in de ToC kunnen we niet met zekerheid stellen dat het door het NGF-project komt, maar als we de beoogde uitkomsten *niet* waarnemen weten we wel dat het NGF-project dit doel niet bereikt heeft. Een eenvoudig voorbeeld hiervan is de technologische voorsprong: het project moet een technologische voorsprong opleveren. Als Nederland die voorsprong straks niet heeft, heeft het NGF-project per definitie niet bijgedragen aan de voorsprong, omdat de voorsprong niet bestaat. Wanneer er wel een voorsprong bestaat zou dit (mede) door het NGF-project kunnen komen, maar het zou ook nog toe te schrijven kunnen zijn aan andere factoren.

Merk op dat **hoe plausibeler de impact pathways in de Theory of Change zijn, hoe aannemelijker het is dat er een causale relatie bestaat wanneer de Hoop-test doorstaan wordt.** Ter illustratie: in de casus "ik drink een kopje koffie, en daardoor wint onze burgemeester een Nobelprijs" is het niet aannemelijk dat er een causale relatie bestaat, zelfs al zou de burgemeester een Nobelprijs winnen en dus de Hoop-test doorstaan. Daarentegen zou in de casus "ik drink een kopje koffie met de burens, en daardoor wordt de band met de burens versterkt" het aannemelijker zijn dat het drinken van de koffie een bijdrage heeft geleverd wanneer we een versterkte band met de burens observeren. Dit illustreert het belang van een goed onderbouwde Theory of Change. Hoe beter onderbouwd de Theory of Change is, hoe relatief eenvoudiger het is (voor evaluatoren) om veronderstelde causale relaties te kunnen onderzoeken, omdat 'minder strenge' testen zoals een Hoop-test voldoende onderbouwing kunnen bieden.

Dit evaluatieontwerp kan verbeterd of 'geüpgraded' worden door betere testen te formuleren. Zo zouden we de verschillende Hoop-testen idealiter upgraden naar Doubly Decisive testen, opdat we geobserveerde positieve uitkomsten met nog meer bewijskracht aan de veronderstelde causale relaties kunnen toeschrijven. Het voorliggende ontwerp kan in deze vorm dus als basis dienen voor de evaluatie, én het is mogelijk om het ontwerp nog verder aan te scherpen.

5.2.5 [4] Accounting for other influencing factors

Een Theory of Change is waardevol voor het onderbouwen, monitoren en evalueren van een NGF-project. **Een inherente zwakte van de Theory of Change is dat deze 'modellering van de werkelijkheid' zich beperkt tot de factoren die door de opstellers geëxpliciteerd zijn.** Het kan immers zo zijn dat er onbenoemde factoren zijn die wel degelijk van invloed zijn op de diverse onderdelen van de Theory of Change, maar waar de opstellers zich niet bewust van waren en/of om andere redenen niet geëxpliciteerd hebben. Zouden we deze onbenoemde factoren negeren in de evaluatie, zou dit ertoe kunnen leiden dat [1] er onterecht veel/weinig van de geobserveerde uitkomsten aan het NGF-project geattributioneerd worden en [2] onbenoemde positieve/negatieve neven-effecten van het NGF-project over het hoofd gezien worden. Om een zo realistisch beeld te vormen van de doeltreffendheid van het NGF-project dient er in de evaluatie expliciet gekeken te worden naar deze onbenoemde 'andere factoren'. De onbenoemde factoren kunnen in twee categorieën opgesplitst worden, te weten [1] onbenoemde gevolgen en [2] onbenoemde oorzaken. Merk op dat ook op dit punt de kwaliteit van de Theory of Change relevant is: een beter onderbouwde Theory of Change zal doorgaans meer oog hebben voor allerlei (externe) factoren en zal derhalve relatief minder relevante factoren onbenoemd laten.

Onbenoemde gevolgen

De **onbenoemde gevolgen** zijn per definitie factoren die beïnvloedbaar zijn, maar zijn niet in de ToC expliciet benoemd. De onbenoemde gevolgen kunnen in iedere laag van de ToC plaatsvinden, van onbenoemde throughputs die het gevolg zijn van de inputs tot aan onbenoemde impacts ten gevolge van de outcomes. In Figuur 13 zijn dit de sterren met nummers 21, 23, 25, 27 en 29. Binnen de evaluatie zal achterhaald moeten worden of er onbenoemde gevolgen zijn en welke dat zijn. We stellen voor dit te doen door:

- **relevante/aanpalende documentatie** te bestuderen, zoals de voortgangsrapportage RVO, expertbeoordelingen van het plan, en sectorrapporten;
- **stakeholders en experts te interviewen** en hen te vragen of de inputs hebben geleid tot andere throughputs die niet in de ToC genoemd zijn, of de throughputs leiden tot andere outputs die niet in de ToC genoemd zijn, etc.

De onbenoemde gevolgen kunnen **wel/niet vooraf bekend zijn bij de evaluatoren**. Indien evaluatoren al concrete ideeën hebben over mogelijke onbenoemde gevolgen, kunnen ze deze concreet toetsen. Dat kan helpen bij het ophalen van informatie. In

deze context is het bijvoorbeeld voor **STER #21** denkbaar dat men binnen de kaders van het NGF-project ook afstemt met de politiek/overheid voor het ontwikkelen van maatwerkafspraken die op hun beurt impact kunnen hebben op het behoud van de staalindustrie in Nederland en de hierop volgende impact op het langetermijnverdienvermogen. Daarnaast kunnen evaluatoren altijd open vragen naar onbenoemde gevolgen die (nog) niet bekend zijn bij hen.

De onbenoemde gevolgen kunnen daarnaast **wel/niet bekend zijn bij geraadpleegde bronnen** (in de vorm van documentatie, gesprekspartners, begeleidingscommissie van een evaluatie). Indien de onbenoemde gevolgen wel bekend zijn bij geraadpleegde bronnen kunnen deze binnen de evaluatie inzichtelijk gemaakt worden. Indien de onbenoemde gevolgen ook bij geraadpleegde bronnen niet naar boven komen, blijven deze onbekend.

Onbenoemde oorzaken

De **onbenoemde oorzaken** kunnen zowel te beïnvloeden zijn als niet te beïnvloeden zijn. Ter illustratie: de investeringen in duurzame energieopwekking zijn waarschijnlijk moeilijk te beïnvloeden door de subsidieaanvragers, maar de parallelle investeringen die nodig zijn om Groen Staal op te kunnen schalen zijn daarentegen juist sterk beïnvloedbaar.

Ook voor de onbenoemde oorzaken geldt dat we willen achterhalen of er onbenoemde oorzaken zijn en welke dat zijn. In Figuur 13 zijn dit de sterren met nummers 20, 22, 24, 26 en 28. Onbenoemde oorzaken kunnen immers een alternatieve verklaring vormen voor de zaken die we wél kunnen observeren. We stellen voor dit te doen door:

- **relevante/aanpalende documentatie te bestuderen**, zoals de voortgangsrapportage RVO, expertbeoordelingen van het plan, en sectorrapporten;
- **stakeholders en experts te interviewen** en hen te vragen of er andere inputs zijn die hebben geleid tot de throughputs in de ToC, andere throughputs die hebben geleid tot de outputs in de ToC, etc.

Ook voor de onbenoemde oorzaken geldt dat deze **wel/niet vooraf bekend kunnen zijn bij de evaluatoren**. Indien evaluatoren al concrete ideeën hebben over mogelijke onbenoemde oorzaken, kunnen ze deze concreet toetsen. Dat kan helpen bij het ophalen van informatie. In deze context is het bijvoorbeeld voor **STER #24** denkbaar dat er andere outputs zijn die leiden tot een technologische voorsprong, zoals parallelle investeringen van Tata Steel Nederland. Daarnaast kunnen evaluatoren altijd open vragen stellen over onbenoemde oorzaken die (nog) niet bekend zijn bij hen.

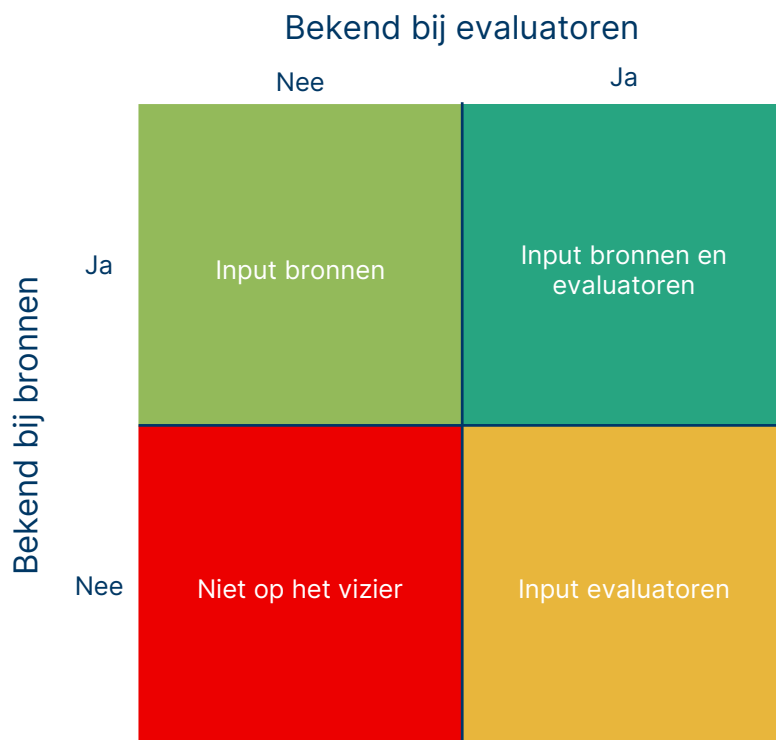
En ook hier geldt dat de onbenoemde oorzaken **wel/niet bekend kunnen zijn bij geraadpleegde bronnen** (in de vorm van documentatie en gesprekspartners). Indien de onbenoemde oorzaken wel bekend zijn bij geraadpleegde bronnen kunnen deze binnen de evaluatie inzichtelijk gemaakt worden. Indien de onbenoemde oorzaken ook bij

geraadpleegde bronnen (documenten en bevroagden) niet naar boven komen, blijven deze onbekend.

Zicht op onbenoemde oorzaken en gevolgen

Indien zowel de evaluatoren als de bronnen (documenten en gesprekspartners) geen zicht bieden op onbenoemde oorzaken en gevolgen, dan zullen deze relevante factoren onopgemerkt blijven in de evaluatie. Dit laatste willen we in de evaluatie vermijden. We kunnen de kans hierop verkleinen door een breed en divers palet van personen te interviewen en goed ingelezen te zijn in het dossier. Desalniettemin is niet met 100% zekerheid uit te sluiten dat er onbenoemde gevolgen onbenoemd blijven.

In Figuur 14 worden de vier scenario's hierbij schematisch weergegeven. Idealiter kunnen de onbenoemde factoren in kaart gebracht worden door zowel de evaluatoren als de bronnen. Een second-best optie is dat het in ieder geval bij de bronnen bekend is. De derde optie is dat het enkel bij de evaluatoren bekend is; een belangrijk aandachtspunt hierbij is dat als dit niet door (enkele) betrokken stakeholders herkend wordt het een speculatief karakter kan krijgen. De vierde optie waarin niemand zicht heeft op relevante externe factoren creëert een blinde vlek in de evaluatie.



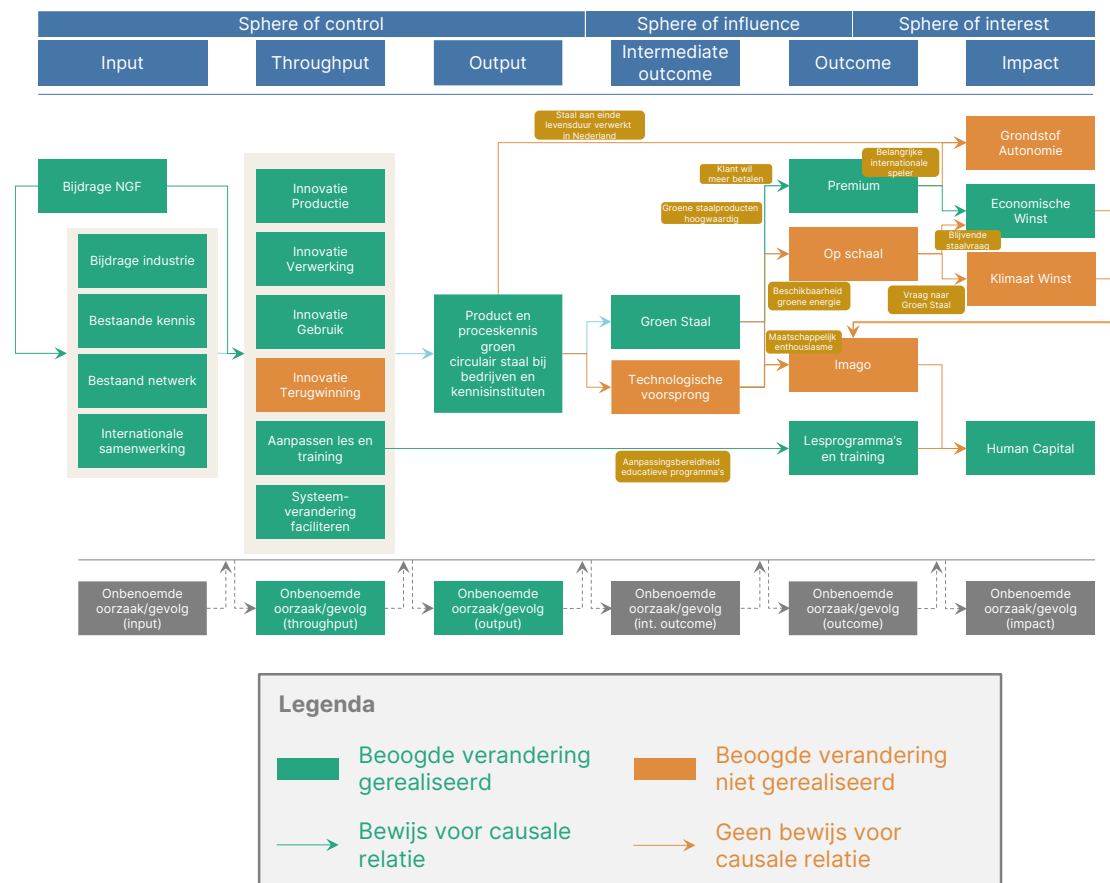
Figuur 14 Bekendheid onbenoemde oorzaken en gevolgen

5.2.6 [5] Synthese

Na het doorlopen van de stappen 1 t/m 4 zoals hiervoor beschreven is de cyclus van de Contribution Analysis doorlopen en kunnen de bevindingen samengebracht worden in deze vijfde en laatste syntheseschap. In deze syntheseschap kunnen alle bevindingen ook visueel samengevat worden door te laten zien waar bewijs is gevonden voor causale relaties in de ToC. Zie Figuur 15 voor een schets van een **fictieve uitkomst** waarin getoond wordt welke onderdelen (blokjes) wel/niet bereikt zijn en voor welke causale mechanismen (pijljes) wel/geen bewijs is gevonden. Op basis van de contribution analysis kan dus aangegeven welke impact pathways wel/niet geslaagd zijn inclusief een toelichting op basis van de verzamelde informatie.

Deze aanpak biedt dus ook de ruimte om een **genuanceerd beeld van de doeltreffendheid** neer te zetten. Het hoeft niet zo te zijn dat het gehele NGF-project wel of niet doeltreffend is geweest. Het kan zijn dat het in termen van bepaalde impact pathways wel doeltreffend is geweest en in termen van andere impact pathways niet. Daar komt ook bij dat er externe factoren zijn die [1] bepalend zijn en [2] niet of nauwelijks te beïnvloeden zijn geweest, zoals de betalingsbereidheid, ontwikkelingen in andere landen en de beschikbaarheid van groene energie. Het kan in theorie zo zijn dat het NGF-project slim is ingericht *gegeven de informatie die men in 2023 had*, maar dat het in 2031 toch niet tot de gewenste resultaten heeft geleid. Het feit dat het project niet doeltreffend zou blijken is niet hetzelfde als zeggen dat het niet slim is geweest om de kans in 2023 te wagen. Dergelijke nuances zijn middels deze aanpak goed aan te brengen, omdat de Theory of Change een basis biedt om deze argumenten gestructureerd in te bedden in de 'logica van het project'.

Om een **empirisch onderbouwde** uitspraak te kunnen doen over de doeltreffendheid van het NGF-project geldt de voorwaarde dat de (intermediate) outcomes en impact een kans moeten hebben gehad om zich te manifesteren. Alleen dan is het mogelijk om een uitspraak te doen over of de beoogde effecten gerealiseerd zijn. Wanneer men eerder in de tijd een uitspraak wil doen over de doeltreffendheid, kan men zich enkel richten op [1] de onderdelen van de Theory of Change die zich tot dan toe hebben kunnen manifesteren in combinatie met [2] een **theoretische onderbouwing** (c.q. een plausibele Theory of Change).



Figuur 15 Fictieve visuele samenvatting doeltreffendheid

5.3 Doelmatigheid GGS

Binnen de 'Regeling periodiek evaluatieonderzoek' wordt de doelmatigheid van beleid gedefinieerd als *"de mate waarin het optimale effect tegen zo min mogelijk kosten en ongewenste neveneffecten wordt bewerkstelligd"*³². Het gaat daarmee in feite om de verhouding tussen de opbrengsten en de kosten. Doelmatigheid wordt ook wel eens beschreven als de doeltreffendheid gedeeld door de kosten. De doeltreffendheid is dan als het ware de teller in de breuk, de kosten de noemer, en doelmatigheid de gehele breuk.

In de praktijk wordt de term doelmatigheid op verschillende manieren benaderd en geanalyseerd. Voor de voorgestelde opzet van de evaluatie van GGS hanteren wij de volgende drie vormen van doelmatigheid, waarbij ze oplopen van een 'nauwe' scope naar een 'brede' scope:

1. **Uitvoeringsdoelmatigheid of operationele doelmatigheid.** Het gaat hier om de doelmatigheid van de uitvoering zelf en bijbehorende kosten, waarbij we kijken naar de volgende drie elementen:

³² <https://wetten.overheid.nl/BWBR0040754/2018-03-27>

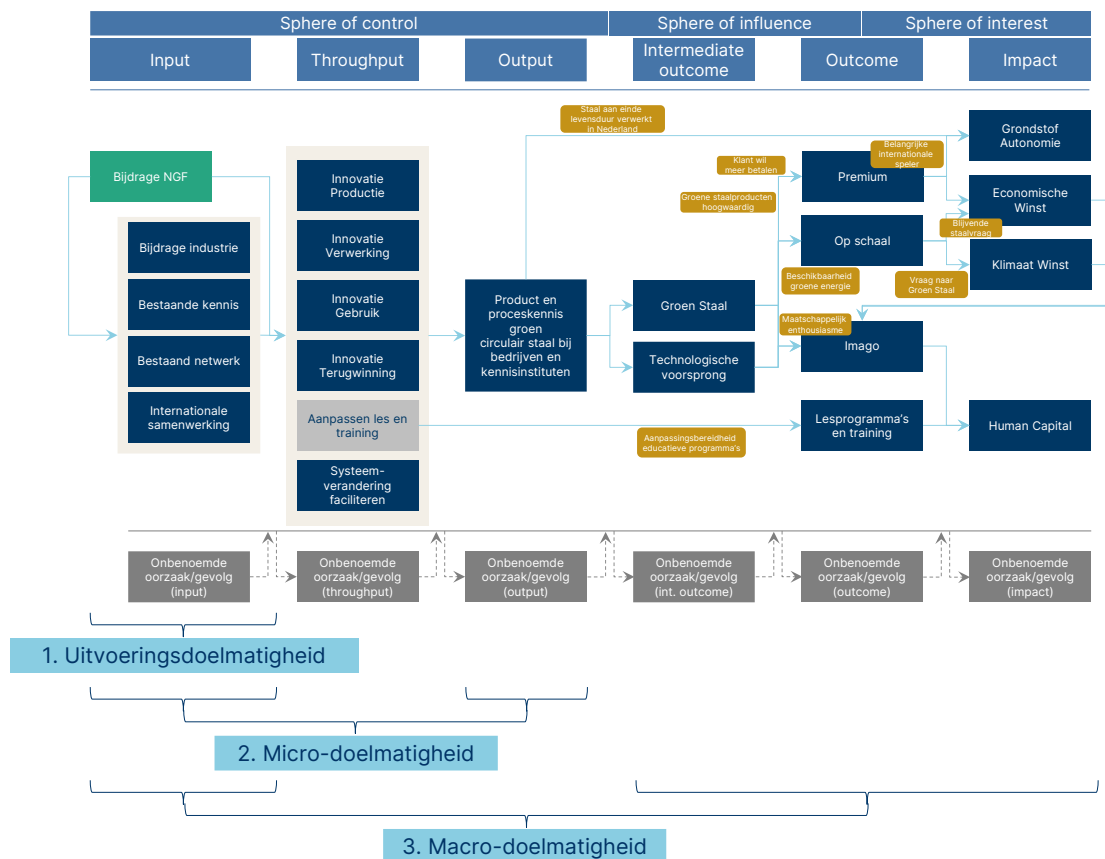
- Het verleende subsidiebedrag. Dit is het bedrag dat als subsidie is uitgekeerd aan de verschillende subsidieontvangers.
- De uitvoering en bijbehorende uitvoeringskosten. De uitvoeringskosten liggen bij degenen die de NGF-subsidie uitvoeren, te weten RVO³³.
- Administratieve lasten. Dit zijn de kosten die subsidieontvangers maken om de subsidie te verkrijgen en te voldoen aan bijbehorende informatieverplichtingen.

Voor de uitvoeringsdoelmatigheid wordt gekeken naar de verhouding tussen de verleende subsidie enerzijds en de uitvoeringskosten plus administratieve lasten anderzijds.

2. **Microdoelmatigheid.** Hiermee doelen we op de opbrengsten op projectniveau (outputs) in verhouding met de gemaakte kosten (subsidiebedrag + uitvoeringskosten + administratieve lasten).
3. **Macrodoelmatigheid.** Hiermee doelen we op de bredere economische en maatschappelijke opbrengsten en neveneffecten (outcomes en impact) in verhouding met de gemaakte kosten (subsidiebedrag + uitvoeringskosten + administratieve lasten)

In Figuur 16 is schematisch weergegeven hoe deze drie vormen van doelmatigheid zich verhouden tot de Theory of Change van GGS.

³³ In NGF-projecten binnen de departementale route kunnen de uitvoeringskosten ook bij departementen en indienende consortia liggen. Voor een indienend consortium gaat het om kosten die zij mogelijk maken om de subsidie uit te zetten; dit is binnen het NGF bijvoorbeeld terug te zien wanneer er stichtingen zijn opgezet die het subsidiegeld verder verdelen of wanneer er calls worden opgezet binnen het NGF-project.



Figuur 16 Uitvoeringsdoelmatigheid, microdoelmatigheid en macrodoelmatigheid

Samengevat zouden we kunnen stellen dat de uitvoeringsdoelmatigheid gaat over de efficiëntie van de subsidieprocessen, de microdoelmatigheid over de projectefficiëntie, en de macrodoelmatigheid over de economische en maatschappelijke efficiëntie van de NGF-subsidie.

De **uitvoeringsdoelmatigheid** kan bepaald worden door de drie onderliggende elementen in kaart te brengen, en vervolgens het percentage uitvoeringskosten plus administratieve lasten als aandeel van de verleende subsidie uit te rekenen. Uit ervaring weten we dat bijvoorbeeld generieke fiscale maatregelen ter bevordering van innovatie doorgaans een percentage van circa 2%-4% kennen, terwijl een instrument als de PPS-toeslag op 6% zat volgens de evaluatie uit 2021. Kleinere regelingen kennen vaak een hoger percentage. Om een oordeel te vellen over de uitvoeringsdoelmatigheid van dit NGF-project dient er een beeld te zijn van wat een 'streefpercentage' zou zijn in deze context van het NGF. Gezien de grote omvang van de subsidiebedragen lijkt het ons reëel om te streven naar een percentage onder de 6% (met de PPS-toeslag als referentie), maar boven de 2% (met de WBSO als referentie). Het verleende subsidiebedrag kan door RVO aangeleverd worden, inclusief de partijen aan wie de subsidie is toegekend. De uitvoeringskosten kunnen geschat worden door voor ieder jaar het aantal fte vast te stellen dat zich heeft gehouden met de subsidie en dat te vermenigvuldigen met het tarief dat de Rijksoverheid op dat moment hanteert om de uitvoeringskosten te berekenen. Dat bedrag is momenteel doorgaans €100.000-

€150.000 per fte. Het aantal fte dient geïnventariseerd te worden bij RVO, EZ en andere departementen die zich beziggehouden hebben met de subsidie. Voor zover er kosten voor de uitvoering worden geoffreerd en afgerekend, heeft het de voorkeur om deze daadwerkelijke bedragen te hanteren in plaats van het gebruiken van een schatting. Voor de uitvoeringskosten die niet berekend kunnen worden op basis van daadwerkelijke bedragen kan de schattingsmethode gebruikt worden. Gegeven dat er twee projecten gesubsidieerd zijn via de subsidieroute van het NGF is het verdedigbaar om de totale uitvoeringskosten m.b.t. de subsidie door tweeën te delen als enkel naar GGS gekeken wordt. De administratieve lasten kunnen in kaart gebracht worden een schatting te maken van de kosten die subsidieontvangers gemaakt hebben om gebruik te kunnen maken van de subsidie. Wanneer de administratieve lasten bij één of enkele partijen neerslaan ligt het voor de hand om 1-op-1 met hen in gesprek te gaan om te achterhalen hoeveel tijd en middelen geïnvesteerd zijn om gebruik te kunnen maken van de subsidie. De in-kind-investeringen in uren kunnen vermenigvuldigd worden met een uurtarief dat plausibel geacht wordt door de betrokkenen. Uurtarieven tussen de €50-€100 lijken realistisch, maar afhankelijk van wie de tijdsinvestering maken kan in de evaluatie een reëel tarief vastgesteld worden. Tot slot kan (op kwalitatieve wijze) inzichtelijk gemaakt worden wat er goed ging in de uitvoering en wat beter zou kunnen. Leer- en verbeterpunten kunnen in gesprek met de betrokken uitvoerders en subsidieontvangers in kaart gebracht worden.

De **microdoelmatigheid** kan bepaald worden door de kosten zoals hierboven beschreven in relatie te zetten tot de opbrengsten op output-niveau. Hierbij kan enerzijds gekeken worden naar de verhouding tussen opbrengsten en kosten en de waardering die verschillende stakeholders hier aan geven. Anderzijds kan bij interviewpartners gericht gevraagd worden naar of deze outputs op een andere manier met minder kosten tot stand hadden kunnen komen. Het uiteindelijke oordeel over de microdoelmatigheid van GGS is deels subjectief, dus het is belangrijk dat de evaluatoren de onderliggende argumenten expliciteren en aangeven hoe zij tot hun oordeel komen. Het expliciet maken van alle kosten en opbrengsten helpt hierbij.

De **macrodoelmatigheid** kan bepaald worden door de kosten zoals hierboven beschreven in relatie te zetten tot de opbrengsten op outcome- en impactniveau. Voor het bepalen van deze vorm van doelmatigheid kan dezelfde werkwijze gehanteerd worden als voor het bepalen van de microdoelmatigheid.

5.4 Vooruitblik

5.4.1 Wat is mogelijk in 2026?

In dit hoofdstuk is een ontwerp voor de evaluatie van de doeltreffendheid en doelmatigheid van GGS gepresenteerd. Dit ontwerp heeft (bewust) nog geen rekening gehouden met de **beperkingen die het moment in tijd van evaluatie** opwerpen. In 2026 is het project pas een paar jaar uit de startblokken en zullen nog niet alle inputs, throughputs en outputs gerealiseerd zijn; ook de hieruit voortvloeiende (intermediate) outcomes en impact zullen daarmee nog niet de kans hebben gehad om zich volledig te manifesteren.

Op basis van het projectplan van GGS kunnen we inschatten welke informatie er in 2026 beschikbaar is. Tabel 8 toont het aantal indicatoren per deel van de Theory of Change dat beschikbaar is in 2026. Alle indicatoren die in 2025 of eerder beschikbaar komen zijn als 'Ja' aangemerkt, alle indicatoren die in 2026 beschikbaar komen zijn als '?' aangemerkt, en alle indicatoren die in 2027 of later beschikbaar komen zijn als 'Nee' aangemerkt. Merk op dat dit overzicht de verwachtingen vanuit het projectplan projetteert; in de praktijk kan het natuurlijk zo zijn dat bepaalde zaken sneller of langzamer verlopen dan verwacht, waardoor bepaalde indicatoren in 2026 al wel of nog niet beschikbaar zijn.

Tabel 8 Beschikbaarheid informatie in 2026. Bron: projectplan GGS, bewerking Dialogic

Beschikbaar in 2026

	Ja	?	Nee	Totaal
Input	5			5
Throughput	27	1		28
Output	18	14	51	83
Int. Outcome		1	7	8
Outcome			19	19
Impact			4	4
Totaal	50	16	81	147

Een belangrijke constatering is dat er **in 2026 (vrijwel) enkel zicht is op de inputs, throughputs en outputs**. Er is naar verwachting nog weinig te zeggen over de intermediate outcomes, outcomes en impact. De tussentijdse evaluatie in 2026 kan wél kijken naar:

- De 'plausibility' van de Theory of Change (criterium 1);
- De 'fidelity': in welke mate zijn de beschreven activiteiten als zodanig uitgevoerd c.q. wat is de projectvoortgang? (criterium 2);
- De eerste impact pathways in de Theory of Change, specifiek sterren 1, 2, 3, en 4 (criterium 3);

- De eerste onbenoemde factoren, specifiek sterren 20, 21, 22 en 23 (criterium 4).

In de tussentijdse evaluatie kan daarnaast wel al met betrokken stakeholders en experts **gesproken worden over de verwachte effecten in de toekomst**. Mogelijk dat er op basis van voortschrijdend inzicht meer onderbouwing te geven is voor waarom bepaalde impact pathways (on)realistischer geworden lijken te zijn.

De evaluatie in 2026 zal hoe dan ook het **karakter krijgen van een tussentijdse evaluatie**, waarbij de nadruk met name zal liggen op de voortgang van het NGF-project, de eerste resultaten en verwachtingen voor de toekomst. Er zal nog geen empirische onderbouwing van de doeltreffendheid en doelmatigheid geboden kunnen worden.

5.4.2 Wat is mogelijk in 2031 en daarna?

In 2031 is er meer mogelijk dan in 2026, omdat er meer informatie over de verschillende onderdelen van de Theory of Change beschikbaar komt. Tegelijkertijd is **de kans klein dat er in 2031 een volledige Contribution Analysis met de voorgestelde process tracing tests uitgevoerd kan worden**, omdat de indicatoren op het gebied van de intermediate outcomes, outcomes en impact pas in 2031 beschikbaar komen (volgens planning). Afhankelijk van wanneer in het jaar de indicatoren beschikbaar komen en wanneer de evaluatie uitgevoerd wordt, kunnen bepaalde indicatoren wel/niet meegenomen worden. In Tabel 9 wordt weergegeven hoeveel indicatoren er in welk jaar beschikbaar komen.

Tabel 9 Beschikbare indicatoren 2031. Bron: projectplan GGS, bewerking Dialogic

Jaar	Input	Throughput	Output	Int. Outcome	Outcome	Impact	Totaal
Lopend	5	27					32
2024			7				7
2025			11				11
2026			14				14
2027			16				16
2028			12				12
2029			15				15
2030			7				7
2031			1	7	13	4	25
2032					1		1
2050					5		5
?		1		1			2
Totaal	5	28	83	8	19	4	147

Hierbij is nog geen rekening gehouden met het feit dat het project vertraging kan oplopen. Gezien de omvang en complexiteit van dergelijke grote NGF-projecten is de kans reëel dat er (op bepaalde onderdelen) vertraging opgelopen zal worden. In dat scenario

is er in 2031 zeker te weinig informatie beschikbaar om een volledige evaluatie uit te kunnen voeren. Op basis van de huidige informatie lijkt het reëler om een *volledige* empirische evaluatie uit te kunnen gaan voeren vanaf 2032 of 2033.

Indien men toch graag eerder in de tijd een evaluatie wil uitvoeren kan men de empirie met de theorie combineren. In een dergelijke situatie kunnen de indicatoren en de process tracing tests tot zover mogelijk uitgevoerd worden ten behoeve van een **empirische onderbouwing** (bijv. tot en met de intermediate outcomes). Voor het resterende 'rechterdeel' in de Theory of Change (outcomes en impact) kan een **theoretische onderbouwing** conform de Theory of Change geboden worden. Voor dit laatste geldt dat de plausibiliteit van (de resterende onderdelen van) de Theory of Change essentieel is. Als het zeer plausibel is, is het voor EZ mogelijk voldoende om de empirische onderbouwing niet voor de gehele Theory of Change op te laten stellen. Is de 'rechterkant van de Theory of Change' echter relatief zwak en niet plausibel, dan volstaat een dergelijke benadering niet om een goed oordeel te kunnen vellen over de doeltreffendheid en doelmatigheid. In een scenario waarin er een evaluatie uitgevoerd dient te worden én nog niet de gehele Theory of Change zich heeft kunnen manifesteren, kan er extra aandacht besteed worden aan het beoordelen van de plausibiliteit, bijvoorbeeld door literatuuronderzoek te doen naar vergelijkbare situaties.

Hoofdstuk 6

Evaluatieontwerp Charging Energy Hubs

In dit hoofdstuk wordt een mogelijk evaluatieontwerp voor NGF-subsidieproject 'Charging Energy Hubs' gepresenteerd.

6 Evaluatieontwerp Charging Energy Hubs

6.1 Inleiding

In dit hoofdstuk presenteren we het evaluatieontwerp voor het bepalen van de doeltreffendheid en doelmatigheid van het NGF-project 'Charging Energy Hubs' (CEH). De eerste evaluatie staat gepland voor 2026. Aangezien het project loopt tot en met 2028, betekent dit dat het project nog niet volledig is uitgevoerd en niet alle beoogde uitkomsten al een kans hebben gehad om zich te manifesteren. Concreet kan in 2026 dus wel wat gezegd worden over de voorlopige resultaten, maar kan nog geen sluitende evaluatie uitgevoerd worden ten aanzien van de doeltreffendheid en doelmatigheid van het gehele NGF-project.

In dit hoofdstuk wordt een evaluatieontwerp voor het gehele NGF-project gepresenteerd. Het hoofdstuk wordt afgesloten met een reflectie op wat naar verwachting reeds mogelijk zal zijn in 2026, en wat mogelijk is in 2028 (einde project) en daarna (0).

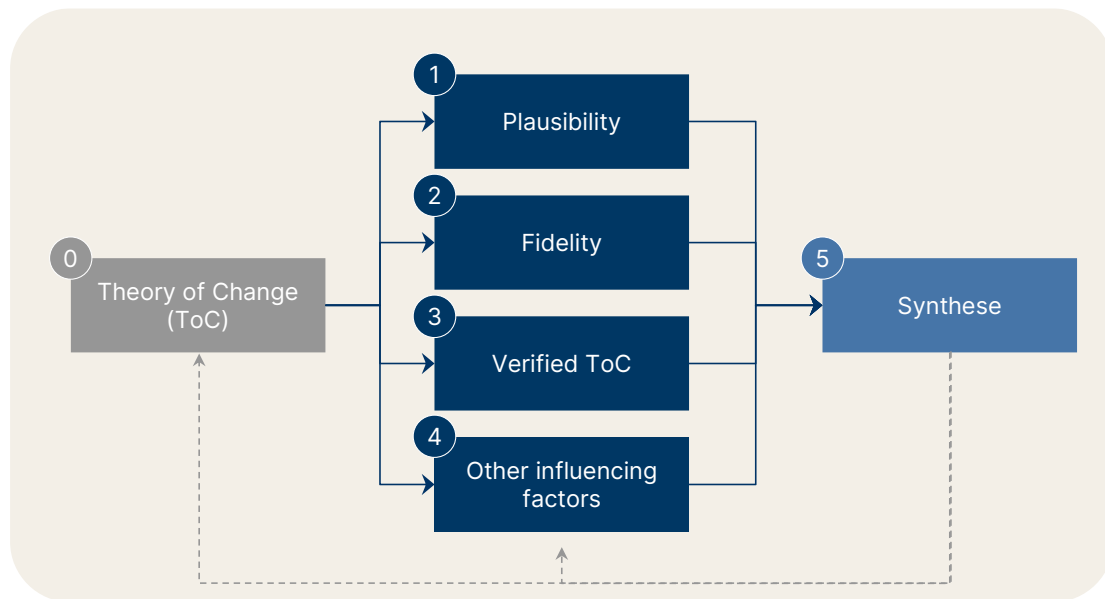
Het evaluatieontwerp wordt opgebouwd aan de hand van de volgende bouwstenen:

Tabel 10: Opzet evaluatieontwerp CEH

Onderdeel evaluatieontwerp	Sectie
Doeltreffendheid CEH	5.2
0. Theory of Change	5.2.1
1. Plausibility	5.2.2
2. Fidelity	5.2.3
3. Verified Theory of Change	5.2.4
4. Other influencing factors	5.2.5
5. Synthese stappen 1 t/m 4	5.2.6
Doelmatigheid	5.3
Vooruitblik	0

6.2 Doeltreffendheid CEH

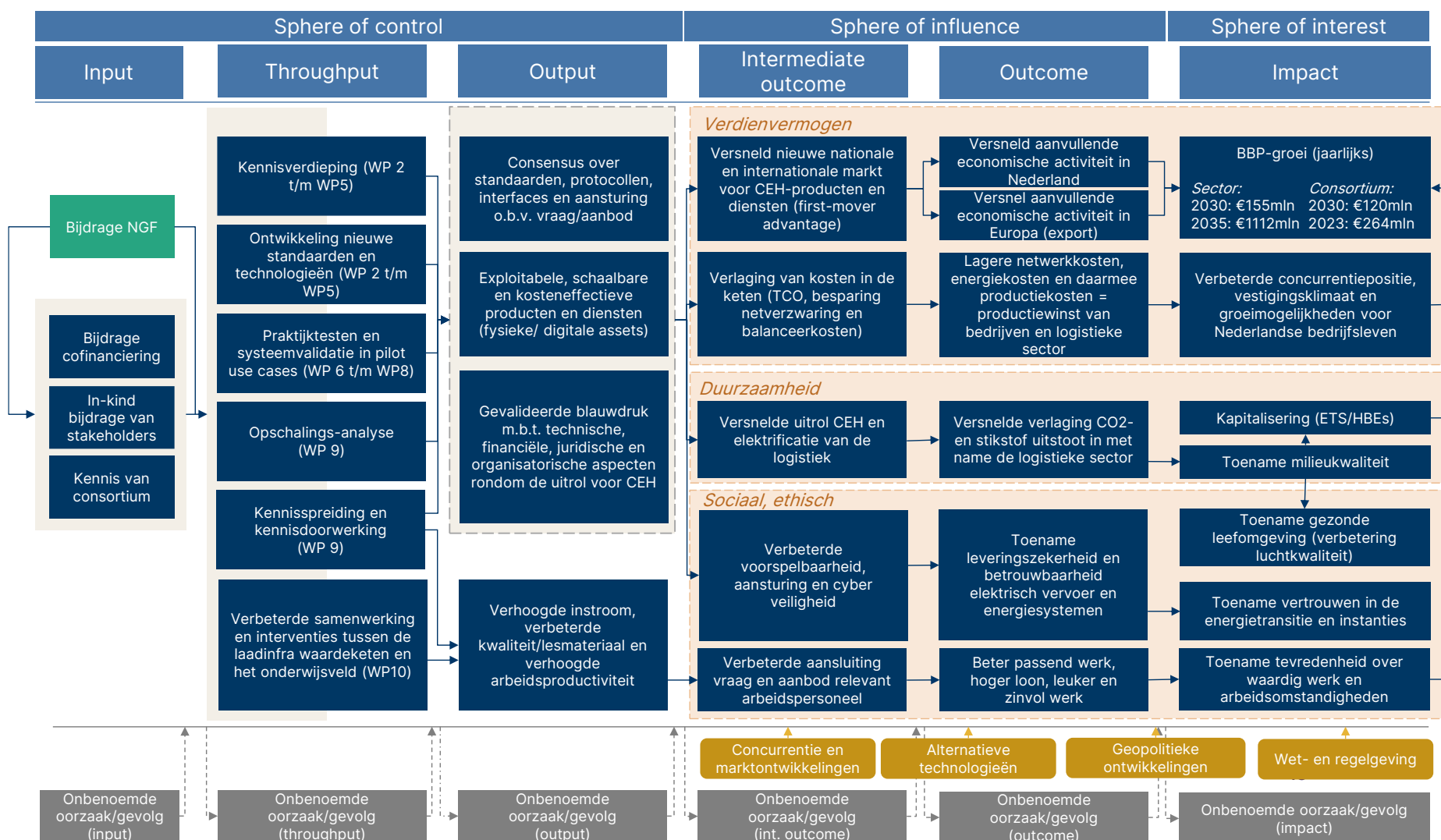
Om de doeltreffendheid van CEH (en de inzet van de NGF-middelen) te beoordelen volgen we de stappen uit de in hoofdstuk 4 beschreven methode:



Figuur 17: Stappen in evaluatiemethode die contributie-analyse en process tracing combineert.

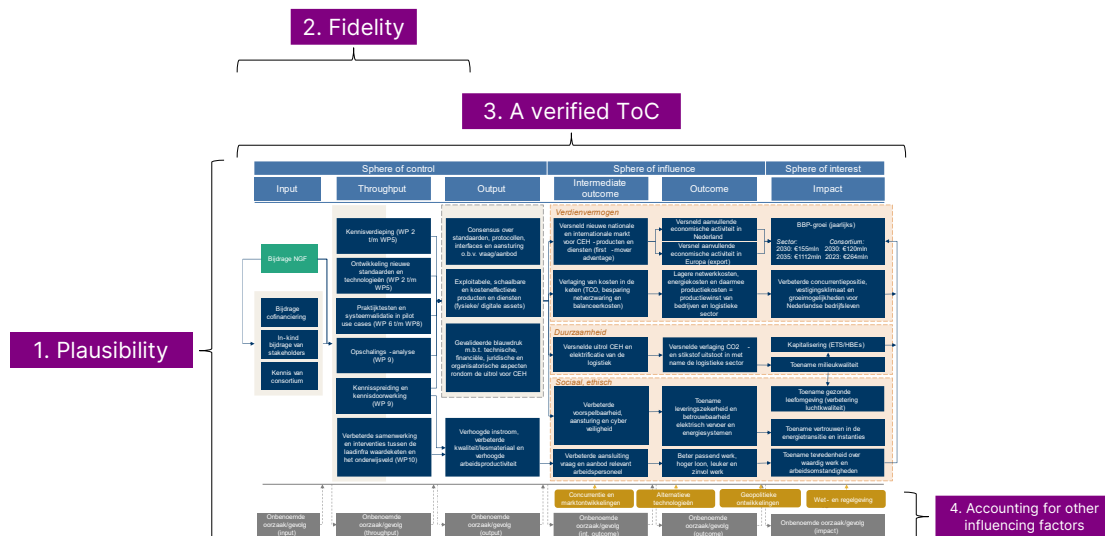
6.2.1 [0] Theory of Change (ToC)

De Theory of Change zoals deze in het projectplan van de subsidieontvangers is gepresenteerd vormt de basis voor de contribution analysis. In Figuur 18 wordt deze Theory of Change weergegeven. Voor meer informatie over de (inhoudelijke) opzet van dit NGF-project verwijzen we terug naar sectie 2.2.



Figuur 18: Theory of Change CEH. Bron: projectplan CEH, bewerking Dialogic

De vier condities waar binnen een contribution analysis aan voldaan moet worden om causaliteit vast te stellen, kunnen als volgt visueel aan de ToC gerelateerd worden:



Figuur 19 De vier criteria geplaatst op de Theory of Change

6.2.2 [1] Plausibility

Het eerste criterium om naar te kijken is 'Plausibility'.

Voor de evaluatie stellen wij voor om in beginsel te **vertrekken vanuit de eerdere expertoordelen** ten aanzien van het NGF-project. Het beoordelen van de plausibiliteit van de Theory of Change vergt immers substantiële expertise op het gebied van de logistieke sector, innovatie en beleid. Hoewel de toekomstige evaluatoren hier potentieel zelf ook een kritische reflectie op kunnen bieden, borgt het aansluiten bij bestaande expertoordelen dat er een solide basis is om de evaluatie goed uit te kunnen voeren.

De centrale bron is de **beoordeling van de Adviescommissie Nationaal Groeifonds**³⁴. De commissie heeft alle inputs van (externe) experts meegenomen en is tot een integraal oordeel gekomen dat in deze bron wordt beschreven. Ten aanzien van de strategische onderbouwing wordt gesteld *"De commissie kan zich goed vinden in de probleemanalyse van het voorstel."* en over de kwaliteit wordt gezegd *"Het plan is helder en in goede mate van detail uitgewerkt en is logisch gestructureerd in verschillende werkpakketten."* De algemene conclusie luidt *"De commissie heeft daardoor (c.q. helder uitgewerkt plan, breed consortium en duidelijke governancestructuur) voldoende vertrouwen dat dit voorstel kan bijdragen aan extra verdienvermogen voor Nederland."* Specifieke aandachtspunten die de commissie benoemt zijn de afstemming met de Autoriteit Consument & Markt (ACM) en de invulling van intellectueel eigendom. Dit zal

³⁴ Commissie Nationaal Groeifonds (juni 2023), Rapport derde beoordelingsronde

nog nader uitgewerkt moeten worden om het succes van het project te kunnen waarborgen.

Een belangrijke bron van input voor de commissie is het **oordeel van het CPB**. Het CPB heeft een analyse uitgevoerd op het project en heeft de resultaten gepubliceerd³⁵. De commissie heeft gereageerd op de aandachtspunten van het CPB. Het CPB is op bepaalde punten minder positief dan de commissie, met name de beperkte bijdrage van het project aan het verminderen van het congestieprobleem op het elektriciteitsnet wordt als grootste knelpunt gezien; *“Het voorstel levert hooguit een oplossing van tijdelijke aard”*. Hiernaast wordt er ook explicieter gewezen op de mogelijke creatie van nieuwe coördinatieproblemen en het beperkte bereik en de beperkte schaal van de use cases; *“Het voorstel kan dus alleen in potentie bijdragen aan verzachten van de knelpunten.”*.

b. Is aannemelijk dat de initiatieven de problemen verhelpen (gegeven de bij ons beschikbare kennis	Overall beeld Deels. Het voorstel kan via innovaties mogelijk een bijdrage leveren aan de verzachting van het hoofdprobleem (te langzame elektrificatie van de logistieke sector). We zien echter drie belangrijke kanttekeningen: • Het voorstel kan zelf leiden tot een nieuw coördinatieprobleem. Om het gebruik van elektriciteit op een bedrijventerrein efficiënt af te
over het ecosysteem)?	stemmen op het aanbod en het lokale elektriciteitsnetwerk, is samenwerking nodig tussen meerdere partijen. • Het voorstel zou een bijdrage te kunnen leveren aan het verminderen van onbalans op het elektriciteitsnet, maar het lijkt niet waarschijnlijk dat door het voorstel substantieel meer bedrijven aangesloten kunnen worden op het huidige elektriciteitsnet. Het voorstel levert hooguit een oplossing van tijdelijke aard. • Een eventuele bijdrage aan elektrificatie in de logistieke sector zal pas na afloop van de subsidieperiode tot stand komen, als innovaties succesvol blijken en door de markt aangeschaft worden. Het voorstel is gericht op innovatieve use cases en beperkt in schaal. Het voorstel kan dus alleen in potentie bijdragen aan verzachten van de knelpunten.

Figuur 20: Koppeling activiteiten aan doelen. Bron: CPB-analyse voorstellen Nationaal Groeifonds

Uitgaande van het oordeel van de commissie kan op dit moment gesteld worden dat de onderbouwing van het plan en daarmee de Theory of Change over het geheel gezien redelijk plausibel is. Er zijn weliswaar bedenkingen over hoe realistisch bepaalde ‘impact pathways’ zijn, maar een evaluatie kan hier juist expliciet aandacht voor hebben. In de beoogde evaluatie kunnen bovenstaande reflecties aangevuld worden met aanvullende (nieuwe) reflecties en bespiegelingen om een oordeel te vellen over de plausibiliteit van de Theory of Change. Daarbij kunnen de evaluatoren de 17 criteria voor het toetsen van de kwaliteit van een Theory of Change meenemen in hun oordeel over de plausibiliteit, zie 4.2.

³⁵ CPB-analyse voorstellen Nationaal Groeifonds (2023), Derde beoordelingsronde, 2023

6.2.3 [2] Fidelity

Het tweede criterium gaat over de '**Fidelity**', ofwel: zijn de acties zoals beschreven in de Theory of Change ook daadwerkelijk als zodanig uitgevoerd?

We baseren ons hier in beginsel op de subsidiebeschikking, het projectplan en de aankomende **voortgangsrapportages** waarin de voortgang op de activiteiten (per werkpakket) worden beschreven. Op basis van de voortgangsrapportages kan vastgesteld worden of de acties inderdaad zijn uitgevoerd zoals beschreven.

In generieke zin geldt dat de evaluatoren oog zullen moeten hebben voor eventuele wijzigingen ten opzichte van het oorspronkelijke projectplan. Deze wijzigingen kunnen op hun beurt ook doorwerken in de rest van de Theory of Change. Het is daarom nuttig om **voorafgaand aan de evaluatie in beeld te hebben wat de wijzigingen zijn en wat de verwachte doorwerkingseffecten hiervan zijn in de Theory of Change**. Als er al duidelijke verwachtingen zijn waarom bepaalde genoemde effecten niet/minder plaats zullen vinden en/of bepaalde nieuwe effecten kunnen gaan optreden, kan de Theory of Change die als basis dient voor het evaluatieontwerp aangepast worden.

6.2.4 [3] Verified Theory of Change

Het derde criterium heeft betrekking op een '**Verified Theory of Change**'.

Voor ieder blokje in de Theory of Change is geïnventariseerd welke data en indicatoren beschikbaar zijn of worden. Het volledige overzicht hiervan is als bijlage beschikbaar voor EZ en RVO, maar kan omwille van vertrouwelijkheid niet gepubliceerd worden. In de tabel hieronder is een overzicht gegeven van het aantal indicatoren per onderdeel in de ToC.

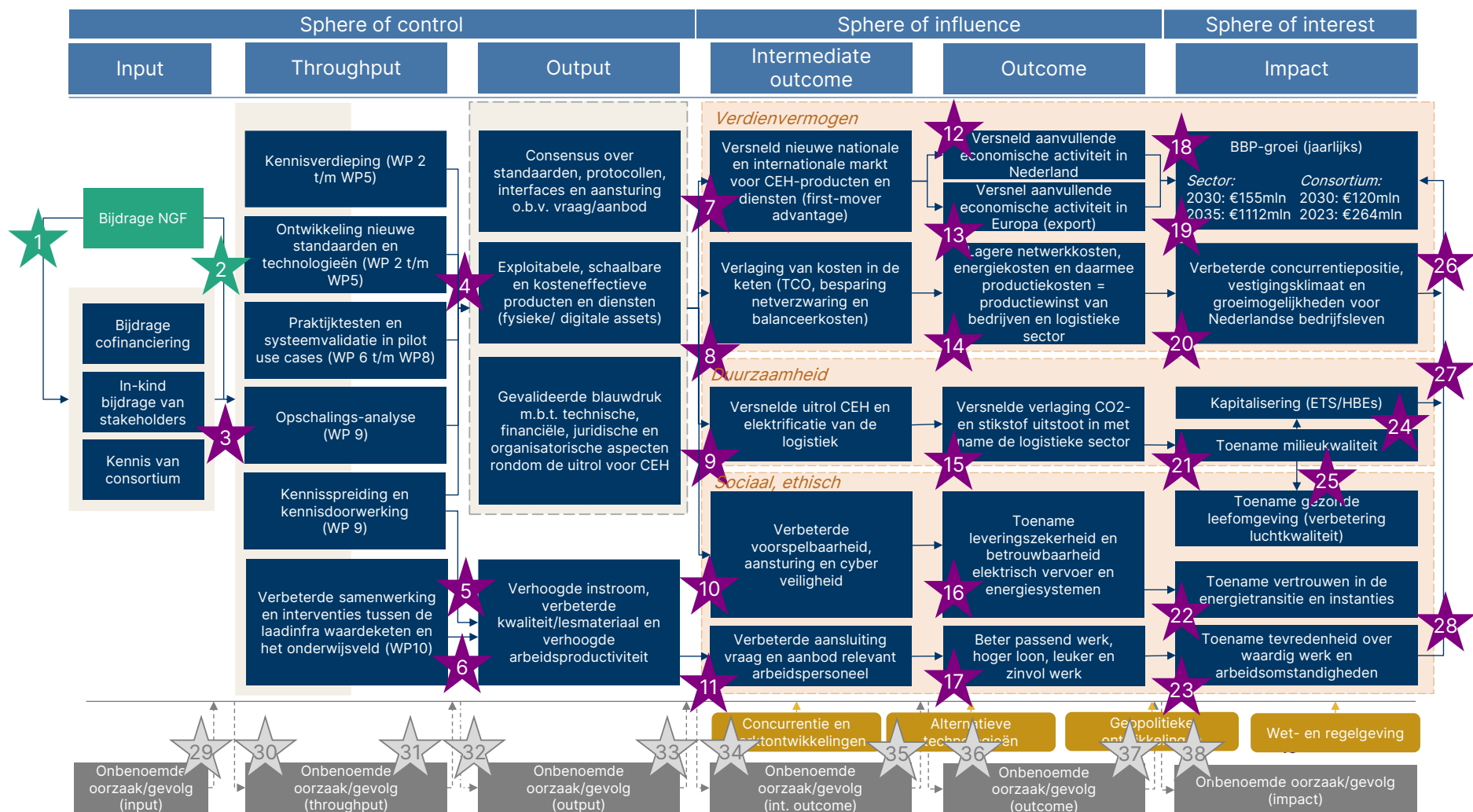
Tabel 11 Aantal beschikbare indicatoren per onderdeel in de ToC

Onderdeel van de ToC	Aantal indicatoren in projectplan	Aantal suggesties ter aanvulling
Input	4	-
Throughput	13	-
Output	231	-
Int. Outcome	4	3
Outcome	1	6
Impact	7	1
Totaal	260	10

Bovenstaande tabel geeft inzicht in het aantal indicatoren per onderdeel van de Theory of Change, maar geeft nog niet weer of voor alle blokjes hierbinnen indicatoren opgesteld zijn. In de onderstaande figuur wordt getoond voor welke blokjes wél indicatoren geformuleerd zijn (**donkerblauw gekleurde blokjes**) en voor welke blokjes géén indicatoren geformuleerd zijn (**lichtblauw gekleurde blokjes**) in het projectplan. Wat direct opvalt is dat er voor een groot aantal intermediate outcomes, outcomes en impacts geen indicatoren zijn opgesteld binnen het projectplan. Wij hebben als onderzoekers enkele eerste suggesties gedaan om de lichtblauw gekleurde blokjes toch meetbaar te maken.

Figuur 21 In het projectplan geformuleerde indicatoren per blokje (donkerblauw = wel geformuleerde indicatoren, lichtblauw = geen geformuleerde indicatoren)

Om op basis van de opgestelde indicatoren iets te zeggen over causale relaties, is het zaak om ook de pijltjes (schakels) in de Theory of Change te onderzoeken en te toetsen. In Figuur 22 worden de pijltjes genummerd (met sterretjes) waarvoor een process tracing test opgesteld is.



Figuur 22 De 'impact pathways' die getest worden in de evaluatie van CEH

Zoals in sectie 3.2 is beschreven zijn er vier typen testen, te weten Straw-in-the-Wind, Hoop, Smoking Gun, en Doubly decisive. Hieronder worden de voorgestelde testen voor CEH beschreven. De nummering is conform de nummering in de sterren in Figuur 22. De beschrijvingen van de testen zijn omwille van vertrouwelijkheid gepseudonimiseerd conform de nummering van de indicatoren die overeenkomen met de bijlage die beschikbaar is voor EZ en RVO.

Tabel 12 Process tracing tests voor CEH

ID ToC	Type test	Type criterium	Beschrijving test	Uitvoering test (dataverzamelmethode en databron)
1	Hoop	Noodzakelijke voorwaarde	Hebben betrokken partijen vanaf de start van het NGF-traject meer tijd en middelen aangewend om te investeren in de uitrol van charging energy hubs? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, dan is de hypothese dat de NGF-bijdrage leidt tot meer input verworpen	- Deskresearch: voortgangsrapportage RVO, projectplan, expertbeoordeling projectplan, jaarverslagen deelnemers project - Interviews: bevragen betrokken stakeholders en (externe) experts over inputadditionaliteit
2	Hoop	Noodzakelijke voorwaarde	Worden er binnen CEH activiteiten ondernomen die zonder NGF-bijdrage niet/minder/anders/later ondernomen zouden zijn? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, dan is de hypothese dat de NGF-bijdrage leidt tot meer/andere throughput verworpen	- Deskresearch: voortgangsrapportage RVO, projectplan, expertbeoordeling projectplan, jaarverslagen deelnemers project - Interviews: bevragen betrokken stakeholders en (externe) experts over inputadditionaliteit
3	Doubly Decisive	Doorslaggevend bewijs	Worden de voorgestelde activiteiten uitgevoerd op basis van de benoemde inputs? - Zo ja, bevestigt hypothese en verworpt hypothese dat activiteiten (enkel) door andere inputs tot stand zijn gekomen - Zo nee, verworpt hypothese dat activiteiten tot stand zijn gekomen (mede) vanwege de benoemde inputs	- Deskresearch: voortgangsrapportage RVO - Interviews: bevragen betrokken stakeholders
4	Doubly Decisive	Doorslaggevend bewijs	Worden de voorgestelde outputs (milestones en deliverables) gecreëerd op basis van de benoemde activiteiten? - Zo ja, bevestigt hypothese en verworpt hypothese dat outputs (enkel) door andere inputs tot stand zijn gekomen - Zo nee, verworpt hypothese dat outputs tot stand zijn gekomen (mede) vanwege de benoemde activiteiten	- Deskresearch: voortgangsrapportage RVO - Interviews: bevragen betrokken stakeholders
5	Doubly Decisive	Doorslaggevend bewijs	Worden de voorgestelde outputs (milestones en deliverables) gecreëerd op basis van de benoemde activiteiten? - Zo ja, bevestigt hypothese en verworpt hypothese dat outputs (enkel) door andere inputs tot stand zijn gekomen - Zo nee, verworpt hypothese dat	- Deskresearch: voortgangsrapportage RVO - Interviews: bevragen betrokken stakeholders

			outputs tot stand zijn gekomen (mede) vanwege de benoemde activiteiten	
6	Doubly Decisive	Doorslaggevend bewijs	Worden de voorgestelde outputs (milestones en deliverables) gecreëerd op basis van de benoemde activiteiten? - Zo ja, bevestigt hypothese en verwerpt hypothese dat outputs (enkel) door andere inputs tot stand zijn gekomen - Zo nee, verwerpt hypothese dat outputs tot stand zijn gekomen (mede) vanwege de benoemde activiteiten	- Deskresearch: voortgangsrapportage RVO - Interviews: bevragen betrokken stakeholders
7a	Doubly Decisive	Doorslaggevend bewijs	[1] Is er sprake van een versneld nieuwe nationale en internationale markt voor CEH-producten en diensten c.q. wordt indicator #250 behaald en [2] zijn hiervoor kennis en/of resultaten gebruikt die enkel binnen/vanwege het NGF-project tot stand zijn gekomen? - Zo ja, bevestigt hypothese en verwerpt hypothese dat een first-mover advantage (enkel) door andere oorzaken tot stand is gekomen. - Zo nee, verwerpt hypothese dat het first-mover advantage tot stand is gekomen vanwege het NGF-project	- Deskresearch: voortgangsrapportage RVO - Interviews: bevragen (externe) experts over outputadditionaliteit
7b	Hoop	Noodzakelijke voorwaarde	Is er sprake van een versneld nieuwe nationale en internationale markt voor CEH-producten en diensten c.q. wordt indicator #250 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, dan is de hypothese dat het NGF-project leidt tot een first-mover advantage verworpen	- Deskresearch: voortgangsrapportage RVO - Interviews: bevragen (externe) experts over outputadditionaliteit - Benchmarking NL t.o.v. andere landen (Zweden, Duitsland, Japan) d.m.v. deskresearch: Marktbeschrijvingen, jaarrapportages
8a	Doubly Decisive	Doorslaggevend bewijs	[1] Vindt er verlaging van de kosten in de logistieke keten plaats c.q. worden indicatoren #251 tot en met #253 behaald en [2] zijn hiervoor kennis en/of resultaten gebruikt die enkel binnen/vanwege het NGF-project tot stand zijn gekomen? - Zo ja, bevestigt hypothese en verwerpt hypothese dat een verlaging van de kosten (enkel) door andere oorzaken tot stand is gekomen. - Zo nee, verwerpt hypothese dat het de lagere kosten tot stand is gekomen vanwege het NGF-project	- Deskresearch: voortgangsrapportage RVO, financiële documentatie stakeholders - Interviews: bevragen betrokken stakeholders

8b	Hoop	Noodzakelijke voorwaarde	Vindt er verlaging van de kosten in de logistieke keten plaats c.q. worden indicatoren #251 tot en met #253 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH tot lagere kosten leidt	- Deskresearch: voortgangsrapportage RVO, financiële documentatie stakeholders - Interviews: bevragen betrokken stakeholders
9a	Doubly Decisive	Doorslaggevend bewijs	[1] Is er sprake van een versnelde uitrol van CEH en elektrificatie van de logistiek c.q. wordt indicator #254 behaald en [2] zijn hiervoor kennis en/of resultaten gebruikt die enkel binnen/vanwege het NGF-project tot stand zijn gekomen? - Zo ja, bevestigt hypothese en verwerpt hypothese dat een versnelde uitrol en elektrificatie (enkel) door andere oorzaken tot stand is gekomen. - Zo nee, verwerpt hypothese dat het de versnelde uitrol en elektrificatie tot stand is gekomen vanwege het NGF-project	- Deskresearch: voortgangsrapportage RVO, rapporten van TNO en RVO over TCO - Interviews: bevragen betrokken stakeholders
9b	Hoop	Noodzakelijke voorwaarde	Is er sprake van een versnelde uitrol van CEH en elektrificatie van de logistiek c.q. wordt indicator #254 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH leidt tot versnelde uitrol en elektrificatie	- Deskresearch: voortgangsrapportage RVO, rapporten van TNO en RVO over TCO - Interviews: bevragen betrokken stakeholders
10	Hoop	Noodzakelijke voorwaarde	Vindt er een verbetering plaats in voorspelbaarheid, aansturing en cyberveiligheid c.q. wordt indicator #255 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH leidt tot verbeterde voorspelbaarheid, aansturing en cyberveiligheid	- Deskresearch: ElaadNL (voorspellingsmodellen), RVO (laadbehoefte), TenneT (aanbod), aantal binnengekomen klachten monitoren
11	Hoop	Noodzakelijke voorwaarde	Vindt er een verbetering plaats in de aansluiting tussen vraag en aanbod van relevant arbeidspersoneel c.q. wordt indicator #256 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH leidt tot verbeterde aansluiting arbeidspersoneel	- Deskresearch: CBS (statistiek aantal vacatures) of UWV (vacatures per sector) - Interviews: bevragen betrokkenen/brancheorganisaties (VNO-NCW)
12	Hoop	Noodzakelijke voorwaarde	Is er een versnelling waar te nemen in aanvullende economische activiteit in Nederland c.q. wordt indicator #257 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH leidt tot meer economische activiteit in Nederland	- CBS-microdata: lijst betrokken bedrijven + ABR + Productiestatistieken

13	Hoop	Noodzakelijke voorwaarde	Is er een versnelling waar te nemen in de aanvullende economische activiteit in Europa c.q. wordt indicator #258 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH leidt tot meer economische activiteit in Europa	- Deskresearch: CBS (export-waarde), OECD (economische analyse voor vergelijking met andere landen), KVK (exporterende bedrijven)
14	Hoop	Noodzakelijke voorwaarde	Zijn de netwerkkosten, en energiekosten en daarmee productiekosten gedaald c.q. is indicator #259 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH leidt tot een productiewinst	- Deskresearch: CBS (productie, kosten), VNO-NCW/TLN (sectorinformatie) - Interviews/focusgroep: betrokken partijen/stakeholders - CBS-microdata: lijst betrokken bedrijven + ABR + Productiestatistieken
15	Hoop	Noodzakelijke voorwaarde	Vindt er een versnelde verlaging CO₂- en stikstof uitstoot in met name de logistieke sector plaats c.q. worden indicatoren #260 en #261 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH versneld tot lagere CO₂- en stikstofuitstoot leidt	- Deskresearch: Voortgangsrapportages RVO, CBS (uitstoot), RIVM (uitstoot)
16	Hoop	Noodzakelijke voorwaarde	Vindt er een toename plaats in leveringszekerheid en betrouwbaarheid van elektrisch vervoer en energiesystemen c.q. wordt indicator #262 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH to een toename van leveringszekerheid en betrouwbaarheid leidt	- Deskresearch: online enquêtes, aantal klachten binnengekomen bij operators, ElaadNL/TenneT laadgegevens
17	Hoop	Noodzakelijke voorwaarde	Is er sprake van een verbetering in passend werk, loon, en leuker en zinvol werk c.q. wordt indicator #263 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH tot beter passend werk, hoger loon, leuker en zinvol werk leidt	- CBS-microdata: lijst bedrijven in sector + ABR + SPOLISBUS + HoogsteOpITab
18	Hoop	Noodzakelijke voorwaarde	Is er een BBP-effect van 1,1 mld p/j gerealiseerd over de gehele sector in 2035 en hebben bedrijven in het project een BBP-effect van 264 mln p/j gerealiseerd c.q. is indicator #264 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH tot een BBP-groei leidt	- Deskresearch: Voortgangsrapportage RVO, BBP-rapportages (nationale rekenkamer, CPB) - Interviews/focusgroep: betrokken partijen/stakeholders - CBS-micodata: lijst van bedrijven binnen de sector + ABR + Productiestatistieken + NFO

19	Hoop	Noodzakelijke voorwaarde	Is er een BBP-effect van 1,1 mld p/j gerealiseerd over de gehele sector in 2035 en hebben bedrijven in het project een BBP-effect van 264 mln p/j gerealiseerd c.q. is indicator #264 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH tot een BBP-groei leidt	- Deskresearch: Voortgangsrapportage RVO, BBP-rapportages (nationale rekenkamer, CPB) - Interviews/focusgroep: betrokken partijen/stakeholders - CBS-micodata: lijst van bedrijven binnen de sector + ABR + Productiestatistieken + NFO
20	Hoop	Noodzakelijke voorwaarde	Is er sprake van een verbetering van de concurrentiepositie, het vestigingsklimaat en de groeimogelijkheden voor Nederlandse bedrijfsleven c.q. is indicator #265 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH tot een productiviteitswinst en daarmee een verbeterde concurrentiepositie, vestigingsklimaat en groeimogelijkheden leidt	Benchmark - Deskresearch: voortgangsrapportage RVO, marktanalyse-rapporten - Interviews: Betrokken partijen/stakeholders en (internationale) experts
21	Hoop	Noodzakelijke voorwaarde	Is er sprake van een toename in milieukwaliteit c.q. worden indicatoren #267 en #268 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH tot een toename van de milieukwaliteit leidt	- Deskresearch: Voortgangsrapportage RVO, publicaties CO2-reductie en materiaalgebruik sector (PBL) - CBS Statline: CO2-emissies per sector - Interviews: Betrokken partijen/stakeholders en experts
22	Hoop	Noodzakelijke voorwaarde	Is er verhoogd vertrouwen in de energietransitie en betrokken instanties c.q. wordt indicator #270 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH tot een toename van vertrouwen in de energietransitie en instanties leidt	- Deskresearch: Opiniepeilingen (Eurobarometer), rapportages vertrouwen over de energietransitie en in overheidsinstanties (EZ) - Enquête: Nieuwe enquête uitzetten onder publiek - CBS Statline: Institutioneel vertrouwen, milieubewustzijn
23	Hoop	Noodzakelijke voorwaarde	Is er sprake van een toename van tevredenheid over waardig werk en arbeidsomstandigheden c.q. wordt indicator #271 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH tot een toename tevredenheid onder werknemers leidt	- Deskresearch: Medewerkerstevredenheidsonderzoeken, interne HR-gegevens, evaluaties werknemers - Benchmarking: Branche- of sectorrapporten - CBS-micodata: lijst betrokken bedrijven + EBB
24	Doubly Decisive	Doorslaggevend bewijs	Vertegenwoordigt de teruggedrongen CO2-uitstoot een economische waarde c.q. wordt indicator #266 behaald? - Zo ja, bevestigt hypothese en verwerpt hypothese dat toename in economische waarde (enkel) door andere oorzaken tot stand is gekomen. - Zo nee, verwerpt hypothese dat het de toename in economische waarde tot stand is gekomen vanwege toename milieukwaliteit	- Deskresearch: Voortgangsrapportage RVO, prijs ETS-rechten (o.a. EEX-publicaties CO2-reductie en materiaalgebruik sector (PBL) - CBS Statline: CO2-emissies per sector - Interviews: Betrokken partijen/stakeholders en experts

25	Hoop	Noodzakelijke voorwaarde	Is er sprake van een toename van een gezonde leefomgeving c.q. wordt indicator #269 behaald? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH tot een verbetering van de luchtkwaliteit leidt	- Deskresearch: luchtkwaliteit; concentraties van vervuilende stoffen (RIVM), luchtmeting over tijd (luchtmeetnet)
26	Hoop	Noodzakelijke voorwaarde	Is er een BBP-effect van 1,1 mld p/j gerealiseerd over de gehele sector in 2035 en hebben bedrijven in het project een BBP-effect van 264 mln p/j gerealiseerd? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH tot een BBP-groei leidt	- Deskresearch: Voortgangsrapportage RVO, BBP-rapportages (nationale rekenkamer, CPB) - Interviews/focusgroep: betrokken partijen/stakeholders - CBS-micodata: lijst van bedrijven binnen de sector + ABR + Productiestatistiek + NFO
27	Hoop	Noodzakelijke voorwaarde	Is er een BBP-effect van 1,1 mld p/j gerealiseerd over de gehele sector in 2035 en hebben bedrijven in het project een BBP-effect van 264 mln p/j gerealiseerd? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH tot een BBP-groei leidt	- Deskresearch: Voortgangsrapportage RVO, BBP-rapportages (nationale rekenkamer, CPB) - Interviews/focusgroep: betrokken partijen/stakeholders - CBS-micodata: lijst van bedrijven binnen de sector + ABR + Productiestatistiek + NFO
28	Hoop	Noodzakelijke voorwaarde	Is er een BBP-effect van 1,1 mld p/j gerealiseerd over de gehele sector in 2035 en hebben bedrijven in het project een BBP-effect van 264 mln p/j gerealiseerd? - Zo ja, bevestigt relevantie van impact pathway maar stelt hem nog niet vast - Zo nee, verwerpt hypothese dat CEH tot een BBP-groei leidt	- Deskresearch: Voortgangsrapportage RVO, BBP-rapportages (nationale rekenkamer, CPB) - Interviews/focusgroep: betrokken partijen/stakeholders - CBS-micodata: lijst van bedrijven binnen de sector + ABR + Productiestatistiek + NFO

Zoals de tabel laat zien kunnen de testen variëren in de manier waarop er invulling aan gegeven wordt in termen van de bijbehorende analysetechniek, dataverzamelingstechniek en databronnen. Enkele voorbeelden ter illustratie zijn:

- Voor de test bij **ster #6** wordt bepaald of de voorgestelde outputs (milestones en deliverables) gecreëerd zijn op basis van de benoemde activiteiten? Een combinatie van het bestuderen van de **voortgangsrapportages** en **interviews** met betrokkenen kan hier licht op werpen.
- Voor de test bij **ster #20** kan gebruikgemaakt worden van **benchmarking** om te bepalen of de Nederlandse sector tot de koplopers op het gebied van CEH binnen Europa behoort? Hiervoor kunnen bronnen binnen het project (bijv. voortgangsrapportage) en bronnen buiten het project (bijv. marktrapportages) ingezet worden.
- Voor de test bij **ster #18** kan binnen de **microdata-omgeving van het CBS** onderzocht worden of er sprake is van een toename in toegevoegde waarde bij

betrokken bedrijven en de sector in brede zin. Hiervoor kan een lijst van relevante organisaties (bijv. o.b.v. KVK-nummer) gecombineerd worden met bestaande registraties binnen het CBS, zoals het Algemene Bedrijvenregister, de Productiestatistieken en de NFO (Financiën van niet-financiële ondernemingen).

In de tabel is te lezen dat **dezelfde test voor verschillende impact pathways** (pijltjes) gebruikt wordt. Dat is het geval bij de pijltjes die naar 'BBP-groei' leiden. Op dit punt testen we met dit ontwerp enkel op de noodzakelijke voorwaarde, omdat het lastig is om vast te stellen hoe de separate elementen bijdragen aan de (mogelijk) geobserveerde groei in toegevoegde waarde. Daarbij is de relatie ook niet overal even helder; zo lijkt er (minimaal) een stap overgeslagen te worden bij de relatie tussen 'Toename tevredenheid over waardig werk en arbeidsomstandigheden' en 'BBP-groei'.

Voor veel impact pathways is een Hoop-test geformuleerd. Met een dergelijke test kan in feite de hypothese dat er wel een causale relatie is enkel gefalsificeerd worden. Simpel gezegd: als we uitkomsten waarnemen zoals beschreven in de ToC kunnen we niet met zekerheid stellen dat het door het NGF-project komt, maar als we de beoogde uitkomsten *niet* waarnemen weten we wel dat het NGF-project dit doel niet bereikt heeft. Een eenvoudig voorbeeld hiervan is de verlaging van de kosten in de keten: het project moet een kostenverlaging opleveren. Als de kostenverlaging straks niet heeft plaatsgevonden heeft het NGF-project per definitie niet bijgedragen aan een kostenverlaging, omdat de kostenverlaging niet bestaat. Wanneer er wel een kostenverlaging bestaat zou dit (mede) door het NGF-project kunnen komen, maar het zou ook nog toe te schrijven kunnen zijn aan andere factoren.

Merk op dat **hoe plausibeler de impact pathways in de Theory of Change zijn, hoe aannemelijker het is dat er een causale relatie bestaat wanneer de Hoop-test doorstaan wordt**. Ter illustratie: in de casus "ik drink een kopje koffie, en daardoor wint onze burgemeester een Nobelprijs" is het niet aannemelijk dat er een causale relatie bestaat, zelfs al zou de burgemeester een Nobelprijs winnen en dus de Hoop-test doorstaan. Daarentegen zou in de casus "ik drink een kopje koffie met de burens, en daardoor wordt de band met de burens versterkt" het aannemelijker zijn dat het drinken van de koffie een bijdrage heeft geleverd wanneer we een versterkte band met de burens observeren. Dit illustreert het belang van een goed onderbouwde Theory of Change. Hoe beter onderbouwd de Theory of Change is, hoe relatief eenvoudiger het is (voor evaluatoren) om veronderstelde causale relaties te kunnen onderzoeken, omdat 'minder strenge' testen voldoende onderbouwing kunnen bieden.

Dit evaluatieontwerp kan verbeterd of 'geüpgraded' worden door betere testen te formuleren. Zo zouden we de verschillende Hoop-testen idealiter upgraden naar Doubly Decisive testen, opdat we geobserveerde positieve uitkomsten met nog meer bewijskracht aan de veronderstelde causale relaties kunnen toeschrijven. Het

voorliggende ontwerp kan in deze vorm dus als basis dienen voor de evaluatie, én het is mogelijk om het ontwerp nog verder aan te scherpen.

6.2.5 [4] Accounting for other influencing factors

Een Theory of Change is waardevol voor het onderbouwen, monitoren en evalueren van een NGF-project. **Een inherente zwakte van de Theory of Change is dat deze 'modellering van de werkelijkheid' zich beperkt tot de factoren die door de opstellers geëxpliciteerd zijn.** Het kan immers zo zijn dat er onbenoemde factoren zijn die wel degelijk van invloed zijn op de diverse onderdelen van de Theory of Change, maar waar de opstellers zich niet bewust van waren en/of om andere redenen niet geëxpliciteerd hebben. Zouden we deze onbenoemde factoren negeren in de evaluatie, zou dit ertoe kunnen leiden dat [1] er onterecht veel/weinig van de geobserveerde uitkomsten aan het NGF-project geattribueerd worden en [2] onbenoemde positieve/negatieve neven-effecten van het NGF-project over het hoofd gezien worden. Om een zo realistisch beeld te vormen van de doeltreffendheid van het NGF-project dient er in de evaluatie expliciet gekeken te worden naar deze onbenoemde 'andere factoren'. De onbenoemde factoren kunnen in twee categorieën opgesplitst worden, te weten [1] onbenoemde gevolgen en [2] onbenoemde oorzaken. Merk op dat ook op dit punt de kwaliteit van de Theory of Change relevant is: een beter onderbouwde Theory of Change zal doorgaans meer oog hebben voor allerlei (externe) factoren en zal derhalve relatief minder relevante factoren onbenoemd laten.

Onbenoemde gevolgen

De **onbenoemde gevolgen** zijn per definitie factoren die beïnvloedbaar zijn, maar zijn niet in de ToC expliciet benoemd. De onbenoemde gevolgen kunnen in iedere laag van de ToC plaatsvinden, van onbenoemde throughputs die het gevolg zijn van de inputs tot aan onbenoemde impacts ten gevolge van de outcomes. In Figuur 22 zijn dit de sterren met nummers 30, 32, 34, 36 en 38. Binnen de evaluatie zal achterhaald moeten worden of er onbenoemde gevolgen zijn en welke dat zijn. We stellen voor dit te doen door:

- **relevante/aanpalende documentatie** te bestuderen, zoals de voortgangsrapportage RVO, expertbeoordelingen van het plan, en sectorrapporten;
- **stakeholders en experts te interviewen** en hen te vragen of de inputs hebben geleid tot andere throughputs die niet in de ToC genoemd zijn, of de throughputs leiden tot andere outputs die niet in de ToC genoemd zijn, etc.

De onbenoemde gevolgen kunnen **wel/niet vooraf bekend zijn bij de evaluatoren**. Indien evaluatoren al concrete ideeën hebben over mogelijke onbenoemde gevolgen, kunnen ze deze concreet toetsen. Dat kan helpen bij het ophalen van informatie. In deze context is het bijvoorbeeld voor **STER #30** denkbaar dat men binnen de kaders van het NGF-project ook afstemt met de Autoriteit Consument & Markt (ACM) over mogelijke ontheffingen die op hun beurt impact hebben op de netwerkkosten en de hierop

volgende impact op het vestigingsklimaat. Daarnaast kunnen evaluatoren altijd open vragen naar onbenoemde gevolgen die (nog) niet bekend zijn bij hen.

De onbenoemde gevolgen kunnen daarnaast **wel/niet bekend zijn bij geraadpleegde bronnen** (in de vorm van documentatie, gesprekspartners, begeleidingscommissie van een evaluatie). Indien de onbenoemde gevolgen wel bekend zijn bij geraadpleegde bronnen kunnen deze binnen de evaluatie inzichtelijk gemaakt worden. Indien de onbenoemde gevolgen ook bij geraadpleegde bronnen niet naar boven komen, blijven deze onbekend.

Onbenoemde oorzaken

De **onbenoemde oorzaken** kunnen zowel te beïnvloeden zijn als niet te beïnvloeden zijn. Ter illustratie: de investeringen in duurzame energieopwekking zijn waarschijnlijk moeilijk te beïnvloeden door de subsidieaanvragers, maar de parallelle investeringen die nodig zijn om Groen Staal op te kunnen schalen zijn daarentegen juist sterk beïnvloedbaar.

Ook voor de onbenoemde oorzaken geldt dat we willen achterhalen of er onbenoemde oorzaken zijn en welke dat zijn. In Figuur 22 zijn dit de sterren met nummers 29, 31, 33, 35 en 37. Onbenoemde oorzaken kunnen immers een alternatieve verklaring vormen voor de zaken die we wél kunnen observeren. We stellen voor dit te doen door:

- **relevante/aanpalende documentatie te bestuderen**, zoals de voortgangsrapportage RVO, expertbeoordelingen van het plan, en sectorrapporten;
- **stakeholders en experts te interviewen** en hen te vragen of er andere inputs zijn die hebben geleid tot de throughputs in de ToC, andere throughputs die hebben geleid tot de outputs in de ToC, etc.

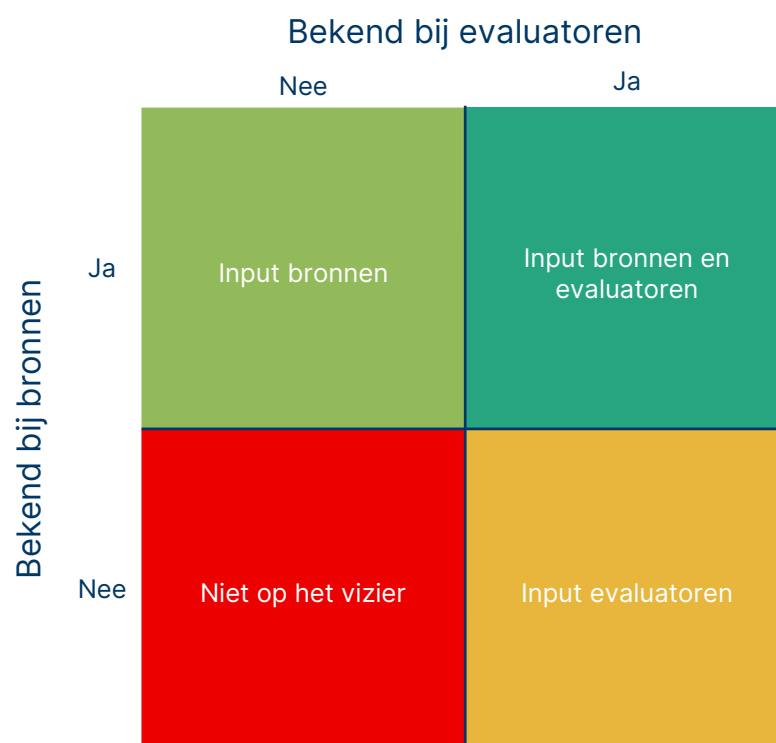
Ook voor de onbenoemde oorzaken geldt dat deze **wel/niet vooraf bekend kunnen zijn bij de evaluatoren**. Indien evaluatoren al concrete ideeën hebben over mogelijke onbenoemde oorzaken, kunnen ze deze concreet toetsen. Dat kan helpen bij het ophalen van informatie. In deze context is het bijvoorbeeld voor **STER #33** denkbaar dat er andere outputs zijn die leiden tot een versnelde elektrificatie van de logistiek, zoals parallelle investeringen in batterijtechnologieën. Daarnaast kunnen evaluatoren altijd open vragen stellen over onbenoemde oorzaken die (nog) niet bekend zijn bij hen.

En ook hier geldt dat de onbenoemde oorzaken **wel/niet bekend kunnen zijn bij geraadpleegde bronnen** (in de vorm van documentatie en gesprekspartners). Indien de onbenoemde oorzaken wel bekend zijn bij geraadpleegde bronnen kunnen deze binnen de evaluatie inzichtelijk gemaakt worden. Indien de onbenoemde oorzaken ook bij geraadpleegde bronnen (documenten en bevrageden) niet naar boven komen, blijven deze onbekend.

Zicht op onbenoemde oorzaken en gevolgen

Indien zowel de evaluatoren als de bronnen (documenten en gesprekspartners) geen zicht bieden op onbenoemde oorzaken en gevolgen, dan zullen deze relevante factoren onopgemerkt blijven in de evaluatie. Dit laatste willen we in de evaluatie vermijden. We kunnen de kans hierop verkleinen door een breed en divers palet van personen te interviewen en goed ingelezen te zijn in het dossier. Desalniettemin is niet met 100% zekerheid uit te sluiten dat er onbenoemde gevolgen onbenoemd blijven.

In Figuur 23 worden de vier scenario's hierbij schematisch weergegeven. Idealiter kunnen de onbenoemde factoren in kaart gebracht worden door zowel de evaluatoren als de bronnen. Een second-best optie is dat het in ieder geval bij de bronnen bekend is. De derde optie is dat het enkel bij de evaluatoren bekend is; een belangrijk aandachtspunt hierbij is dat als dit niet door (enkele) betrokken stakeholders herkend wordt het een speculatief karakter kan krijgen. De vierde optie waarin niemand zicht heeft op relevante externe factoren creëert een blinde vlek in de evaluatie.



Figuur 23 Bekendheid onbenoemde oorzaken en gevolgen

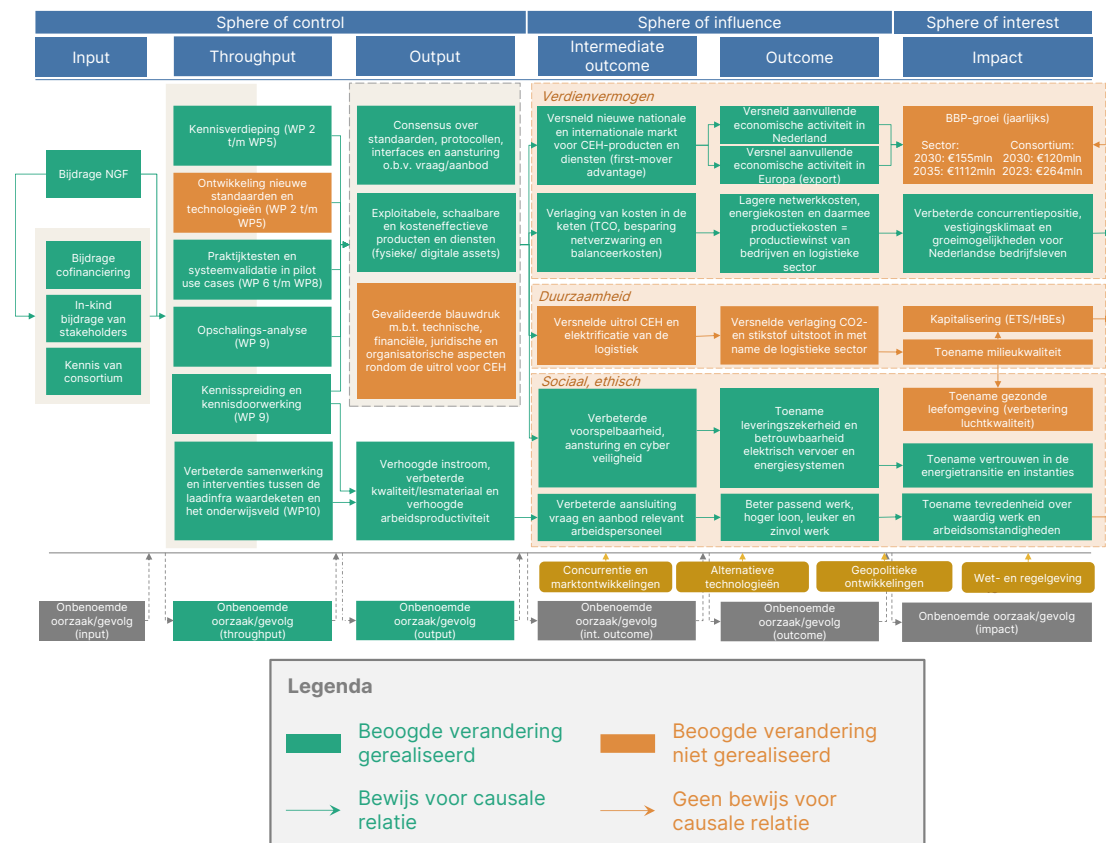
6.2.6 [5] Synthese

Na het doorlopen van de stappen 1 t/m 4 zoals hiervoor beschreven is de cyclus van de Contribution Analysis doorlopen en kunnen de bevindingen samengebracht worden in deze vijfde en laatste syntheses stap. In deze syntheses stap kunnen alle bevindingen ook visueel samengevat worden door te laten zien waar bewijs is gevonden voor

causale relaties in de ToC. Zie Figuur 24 voor een schets van een **fictieve uitkomst** waarin getoond wordt welke onderdelen (blokjes) wel/niet bereikt zijn en voor welke causale mechanismen (pijl-tjes) wel/geen bewijs is gevonden. Op basis van de contribution analysis kan dus aangegeven welke impact pathways wel/niet geslaagd zijn inclusief een toelichting op basis van de verzamelde informatie.

Deze aanpak biedt dus ook de ruimte om een **genuanceerd beeld van de doeltreffendheid** neer te zetten. Het hoeft niet zo te zijn dat het gehele NGF-project wel of niet doeltreffend is geweest. Het kan zijn dat het in termen van bepaalde impact pathways wel doeltreffend is geweest en in termen van andere impact pathways niet. Daar komt ook bij dat er externe factoren zijn die [1] bepalend zijn en [2] niet of nauwelijks te beïnvloeden zijn geweest, zoals de betalingsbereidheid, ontwikkelingen in andere landen en de beschikbaarheid van groene energie. Het kan in theorie zo zijn dat het NGF-project slim is ingericht *gegeven de informatie die men in 2023 had*, maar dat het in 2028 toch niet tot de gewenste resultaten heeft geleid. Het feit dat het project niet doeltreffend zou blijken is niet hetzelfde als zeggen dat het niet slim is geweest om de kans in 2023 te wagen. Dergelijke nuances zijn middels deze aanpak goed aan te brengen, omdat de Theory of Change een basis biedt om deze argumenten gestructureerd in te bedden in de 'logica van het project'.

Om een **empirisch onderbouwde** uitspraak te kunnen doen over de doeltreffendheid van het NGF-project geldt de voorwaarde dat de (intermediate) outcomes en impact een kans moeten hebben gehad om zich te manifesteren. Alleen dan is het mogelijk om een uitspraak te doen over of de beoogde effecten gerealiseerd zijn. Wanneer men eerder in de tijd een uitspraak wil doen over de doeltreffendheid, kan men zich enkel richten op [1] de onderdelen van de Theory of Change die zich tot dan toe hebben kunnen manifesteren in combinatie met [2] een **theoretische onderbouwing** (c.q. een plausibele Theory of Change).



Figuur 24 Fictieve visuele samenvatting doeltreffendheid

6.3 Doelmatigheid CEH

Binnen de 'Regeling periodiek evaluatieonderzoek' wordt de doelmatigheid van beleid gedefinieerd als *"de mate waarin het optimale effect tegen zo min mogelijk kosten en ongewenste neveneffecten wordt bewerkstelligd"*³⁶. Het gaat daarmee in feite om de verhouding tussen de opbrengsten en de kosten. Doelmatigheid wordt ook wel eens beschreven als de doeltreffendheid gedeeld door de kosten. De doeltreffendheid is dan als het ware de teller in de breuk, de kosten de noemer, en doelmatigheid de gehele breuk.

In de praktijk wordt de term doelmatigheid op verschillende manieren benaderd en geanalyseerd. Voor de voorgestelde opzet van de evaluatie van CEH hanteren wij de volgende drie vormen van doelmatigheid, waarbij ze oplopen van een 'nauwe' scope naar een 'brede' scope:

1. **Uitvoeringsdoelmatigheid of operationele doelmatigheid.** Het gaat hier om de doelmatigheid van de uitvoering zelf en bijbehorende kosten, waarbij we kijken naar de volgende drie elementen:

³⁶ <https://wetten.overheid.nl/BWBR0040754/2018-03-27>

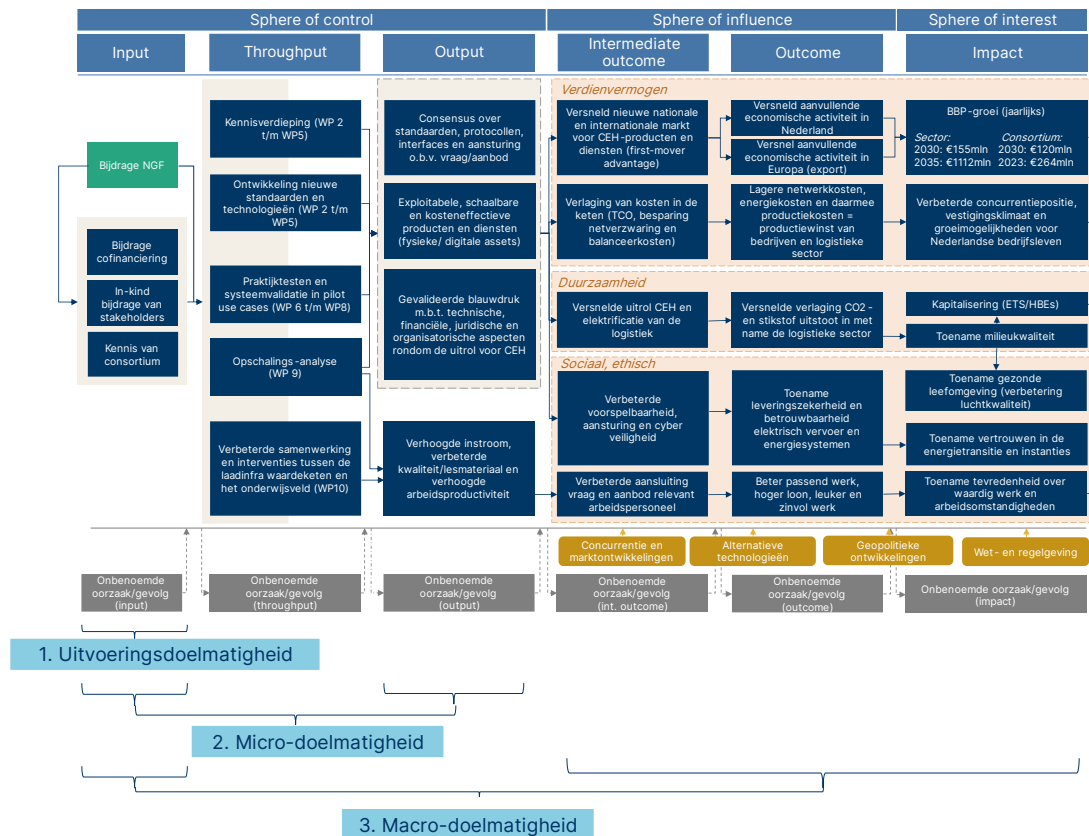
- Het verleende subsidiebedrag. Dit is het bedrag dat als subsidie is uitgekeerd aan de verschillende subsidieontvangers.
- De uitvoering en bijbehorende uitvoeringskosten. De uitvoeringskosten liggen bij degenen die de NGF-subsidie uitvoeren, te weten RVO³⁷.
- Administratieve lasten. Dit zijn de kosten die subsidieontvangers maken om de subsidie te verkrijgen en te voldoen aan bijbehorende informatieverplichtingen.

Voor de uitvoeringsdoelmatigheid wordt gekeken naar de verhouding tussen de verleende subsidie enerzijds en de uitvoeringskosten plus administratieve lasten anderzijds.

2. **Microdoelmatigheid.** Hiermee doelen we op de opbrengsten op projectniveau (outputs) in verhouding met de gemaakte kosten (subsidiebedrag + uitvoeringskosten + administratieve lasten).
3. **Macrodoelmatigheid.** Hiermee doelen we op de bredere economische en maatschappelijke opbrengsten en neveneffecten (outcomes en impact) in verhouding met de gemaakte kosten (subsidiebedrag + uitvoeringskosten + administratieve lasten)

In Figuur 25 is schematisch weergegeven hoe deze drie vormen van doelmatigheid zich verhouden tot de Theory of Change van CEH.

³⁷ In NGF-projecten binnen de departementale route kunnen de uitvoeringskosten ook bij departementen en indienende consortia liggen. Voor een indienend consortium gaat het om kosten die zij mogelijk maken om de subsidie uit te zetten; dit is binnen het NGF bijvoorbeeld terug te zien wanneer er stichtingen zijn opgezet die het subsidiegeld verder verdelen of wanneer er calls worden opgezet binnen het NGF-project.



Figuur 25 Uitvoeringsdoelmatigheid, microdoelmatigheid en macrodoelmatigheid

Samengevat zouden we kunnen stellen dat de uitvoeringsdoelmatigheid gaat over de efficiëntie van de subsidieprocessen, de microdoelmatigheid over de projectefficiëntie, en de macrodoelmatigheid over de economische en maatschappelijke efficiëntie van de NGF-subsidie.

De **uitvoeringsdoelmatigheid** kan bepaald worden door de drie onderliggende elementen in kaart te brengen, en vervolgens het percentage uitvoeringskosten plus administratieve lasten als aandeel van de verleende subsidie uit te rekenen. Uit ervaring weten we dat bijvoorbeeld generieke fiscale maatregelen ter bevordering van innovatie doorgaans een percentage van circa 2%-4% kennen, terwijl een instrument als de PPS-toeslag op 6% zat volgens de evaluatie uit 2021. Kleinere regelingen kennen vaak een hoger percentage. Om een oordeel te vellen over de uitvoeringsdoelmatigheid van dit NGF-project dient er een beeld te zijn van wat een 'streefpercentage' zou zijn in deze context van het NGF. Gezien de grote omvang van de subsidiebedragen lijkt het ons reëel om te streven naar een percentage onder de 6% (met de PPS-toeslag als referentie), maar boven de 2% (met de WBSO als referentie). Het verleende subsidiebedrag kan door RVO aangeleverd worden, inclusief de partijen aan wie de subsidie is toegekend. De uitvoeringskosten kunnen geschat worden door voor ieder jaar het aantal fte vast te stellen dat zich heeft gehouden met de subsidie en dat te vermenigvuldigen met het tarief dat de Rijksoverheid op dat moment hanteert om de uitvoeringskosten te berekenen. Dat bedrag is momenteel doorgaans €100.000-

€150.000 per fte. Het aantal fte dient geïnventariseerd te worden bij RVO, EZ en andere departementen die zich beziggehouden hebben met de subsidie. Voor zover er kosten voor de uitvoering worden geoffreerd en afgerekend, heeft het de voorkeur om deze daadwerkelijke bedragen te hanteren in plaats van het gebruiken van een schatting. Voor de uitvoeringskosten die niet berekend kunnen worden op basis van daadwerkelijke bedragen kan de schattingsmethode gebruikt worden. Gegeven dat er twee projecten gesubsidieerd zijn via de subsidieroute van het NGF is het verdedigbaar om de totale uitvoeringskosten m.b.t. de subsidie door tweeën te delen als enkel naar CEH gekeken wordt. De administratieve lasten kunnen in kaart gebracht worden een schatting te maken van de kosten die subsidieontvangers gemaakt hebben om gebruik te kunnen maken van de subsidie. Wanneer de administratieve lasten bij één of enkele partijen neerslaan ligt het voor de hand om 1-op-1 met hen in gesprek te gaan om te achterhalen hoeveel tijd en middelen geïnvesteerd zijn om gebruik te kunnen maken van de subsidie. De in-kind-investeringen in uren kunnen vermenigvuldigd worden met een uurtarief dat plausibel geacht wordt door de betrokkenen. Uurtarieven tussen de €50-€100 lijken realistisch, maar afhankelijk van wie de tijdsinvestering maken kan in de evaluatie een reëel tarief vastgesteld worden. Tot slot kan (op kwalitatieve wijze) inzichtelijk gemaakt worden wat er goed ging in de uitvoering en wat beter zou kunnen. Leer- en verbeterpunten kunnen in gesprek met de betrokken uitvoerders en subsidieontvangers in kaart gebracht worden.

De **microdoelmatigheid** kan bepaald worden door de kosten zoals hierboven beschreven in relatie te zetten tot de opbrengsten op output-niveau. Hierbij kan enerzijds gekeken worden naar de verhouding tussen opbrengsten en kosten en de waardering die verschillende stakeholders hier aan geven. Anderzijds kan bij interviewpartners gericht gevraagd worden naar of deze outputs op een andere manier met minder kosten tot stand hadden kunnen komen. Het uiteindelijke oordeel over de microdoelmatigheid van CEH is deels subjectief, dus het is belangrijk dat de evaluatoren de onderliggende argumenten expliciteren en aangeven hoe zij tot hun oordeel komen. Het expliciet maken van alle kosten en opbrengsten helpt hierbij.

De **macrodoelmatigheid** kan bepaald worden door de kosten zoals hierboven beschreven in relatie te zetten tot de opbrengsten op outcome- en impactniveau. Voor het bepalen van deze vorm van doelmatigheid kan dezelfde werkwijze gehanteerd worden als voor het bepalen van de microdoelmatigheid.

6.4 Vooruitblik

6.4.1 Wat is mogelijk in 2026?

In dit hoofdstuk is een ontwerp voor de evaluatie van de doeltreffendheid en doelmatigheid van CEH gepresenteerd. Dit ontwerp heeft (bewust) nog geen rekening gehouden met de **beperkingen die het moment in tijd van evaluatie** opwerpen. In 2026 is het project pas een paar jaar uit de startblokken en zullen nog niet alle inputs, throughputs en outputs gerealiseerd zijn; ook de hieruit voortvloeiende (intermediate) outcomes en impact zullen daarmee nog niet de kans hebben gehad om zich volledig te manifesteren.

Op basis van het projectplan van CEH kunnen we inschatten welke informatie er in 2026 beschikbaar is. Tabel 13 toont het aantal indicatoren per deel van de Theory of Change dat beschikbaar is in 2026. Alle indicatoren die in 2025 of eerder beschikbaar komen zijn als 'Ja' aangemerkt, alle indicatoren die in 2026 beschikbaar komen zijn als '?' aangemerkt, en alle indicatoren die in 2027 of later beschikbaar komen zijn als 'Nee' aangemerkt. Merk op dat dit overzicht de verwachtingen vanuit het projectplan projecteert; in de praktijk kan het natuurlijk zo zijn dat bepaalde zaken sneller of langzamer verlopen dan verwacht, waardoor bepaalde indicatoren in 2026 al wel of nog niet beschikbaar zijn.

Tabel 13 Beschikbaarheid informatie in 2026. Bron: projectplan CEH, bewerking Dialogic

Beschikbaar in 2026

	Ja	?	Nee	Totaal
Input	4			4
Throughput	13			13
Output	138	3	90	231
Int. Outcome		3	4	7
Outcome		5	2	7
Impact		1	7	8
Totaal	155	12	103	270

Een belangrijke constatering is dat er **in 2026 (vrijwel) enkel zicht is op de inputs, throughputs en outputs**. Er is naar verwachting nog weinig te zeggen over de intermediate outcomes, outcomes en impact. De tussentijdse evaluatie in 2026 kan wél kijken naar:

- De 'plausibility' van de Theory of Change (criterium 1);
- De 'fidelity': in welke mate zijn de beschreven activiteiten als zodanig uitgevoerd c.q. wat is de projectvoortgang? (criterium 2);
- De eerste impact pathways in de Theory of Change, specifiek sterren 1, 2, 3, en 4 (criterium 3);
- De eerste onbenoemde factoren, specifiek sterren 28, 29, 30 en 31 (criterium 4).

In de tussentijdse evaluatie kan daarnaast wel al met betrokken stakeholders en experts **gesproken worden over de verwachte effecten in de toekomst**. Mogelijk dat er op basis van voortschrijdend inzicht meer onderbouwing te geven is voor waarom bepaalde impact pathways (on)realistischer geworden lijken te zijn.

De evaluatie in 2026 zal hoe dan ook het **karakter krijgen van een tussentijdse evaluatie**, waarbij de nadruk met name zal liggen op de voortgang van het NGF-project, de eerste resultaten en verwachtingen voor de toekomst. Er zal nog geen empirische onderbouwing van de doeltreffendheid en doelmatigheid geboden kunnen worden.

6.4.2 Wat is mogelijk in 2028 en daarna?

In 2028 is er meer mogelijk dan in 2026, omdat er meer informatie over de verschillende onderdelen van de Theory of Change beschikbaar komt. Tegelijkertijd is **de kans klein dat er in 2028 een volledige Contribution Analysis met de voorgestelde process tracing tests uitgevoerd kan worden**, omdat [1] de indicatoren op het gebied van de intermediate outcomes, outcomes en impact grotendeels pas na 2028 beschikbaar komen (volgens planning) en [2] veel outcomes en impacts afhankelijk zijn van indicatoren die niet in het projectplan opgenomen zijn om te meten. In Tabel 14 wordt weergegeven hoeveel indicatoren er in welk jaar beschikbaar komen, waarbij de meeste ‘?’ die genoemd zijn in Figuur 15 overeenkomen met suggesties waarvan niet bekend is of die überhaupt gemeten gaan worden.

Tabel 14 Beschikbare indicatoren 2028. Bron: projectplan CEH, bewerking Dialogic

Jaar	Input	Throughput	Output	Int. Outcome	Outcome	Impact	Totaal
Lopend	4	13					17
2024			51				51
2025			53				53
2026			38				38
2027			86	3			89
2030						1	1
2035				1	2	6	9
?			3	3	5	1	12
Totaal	4	13	231	7	7	8	270

Hierbij is nog geen rekening gehouden met het feit dat het project vertraging kan oplopen. Gezien de omvang en complexiteit van dergelijke grote NGF-projecten is de kans reëel dat er (op bepaalde onderdelen) vertraging opgelopen zal worden.

Op basis van de huidige informatie kunnen we stellen dat er in ieder geval aanvullende inspanningen gedaan moeten worden m.b.t. het ontwikkelen en verzamelen van indicatoren om een volledige evaluatie van alle impact pathways uit te kunnen voeren. Deze inspanningen zijn ook nodig om goed vast te kunnen stellen in welke jaren welke

indicatoren beschikbaar kunnen komen. Daarnaast lijken de impacts en een deel van de outcomes vanaf 2035/2036 inzichtelijk te kunnen worden gemaakt op basis van indicatoren die wel opgenomen zijn in het projectplan. Gegeven de stand van zaken nu is het denkbaar dat een volledige evaluatie van de gehele Theory of Change pas uitgevoerd kan worden na 2035. Dit is in lijn met het CPB-advies³⁸, dat aangaf dat de bijdrage aan elektrificatie in de logistieke sector pas na afloop van de subsidieperiode tot stand zal komen. Dit komt mede door de beperkte schaal van de use cases; als blijkt dat de innovaties succesvol zijn zullen deze eerst door de markt aangeschaft moeten worden wil de impact te zien zijn op grotere schaal.

Indien men toch een evaluatie wil uitvoeren zonder dat de volledige Theory of Change zich heeft kunnen manifesteren, kan men de empirie met de theorie combineren. In een dergelijke situatie kunnen de indicatoren en de process tracing tests tot zover mogelijk uitgevoerd worden ten behoeve van een **empirische onderbouwing** (bijv. tot en met de intermediate outcomes). Voor het resterende 'rechterdeel' in de Theory of Change (outcomes en impact) kan een **theoretische onderbouwing** conform de Theory of Change geboden worden. Voor dit laatste geldt dat de plausibiliteit van (de resterende onderdelen van) de Theory of Change essentieel is. Als het zeer plausibel is, is het voor EZ mogelijk voldoende om de empirische onderbouwing niet voor de gehele Theory of Change op te laten stellen. Is de 'rechterkant van de Theory of Change' echter relatief zwak en niet plausibel, dan volstaat een dergelijke benadering niet om een goed oordeel te kunnen vellen over de doeltreffendheid en doelmatigheid. In een scenario waarin er een evaluatie uitgevoerd dient te worden én nog niet de gehele Theory of Change zich heeft kunnen manifesteren, kan er extra aandacht besteed worden aan het beoordelen van de plausibiliteit, bijvoorbeeld door literatuuronderzoek te doen naar vergelijkbare situaties.

³⁸ CPB-analyse voorstellen Nationaal Groeifonds (2023), Derde beoordelingsronde, 2023

Hoofdstuk 7

Aandachtspunten en aanbevelingen voor evalua- ties van NGF-projecten

In dit hoofdstuk worden overkoepelende aandachtspunten en aanbevelingen voor toekomstige evaluaties van NGF-projecten gepresenteerd.

7 Aandachtspunten & aanbevelingen

In de voorgaande stukken zijn een aantal belangrijke evaluatiebegrippen en -methoden beschreven (hoofdstuk 3), is een voorstel gedaan voor een relatief universele methode voor het evalueren van NGF-projecten (hoofdstuk 4), en is deze methode uitgewerkt voor de twee NGF-subsidieprojecten Groeien met Groen Staal en Charging Energy Hubs (hoofdstukken 5 en 6). Tijdens dit vooronderzoek en het uitwerken van de evaluatieontwerpen zijn een aantal aandachtspunten naar voren gekomen die relevant zijn voor de evaluatie van de twee specifieke subsidieprojecten, maar ook voor de toekomstige evaluaties van de departementale NGF-projecten. In dit hoofdstuk staan we stil bij deze aandachtspunten en formuleren we enkele aanbevelingen om constructief met deze aandachtspunten om te kunnen gaan.

We staan eerst stil bij de voorgestelde evaluatiemethode in zijn geheel (7.1) en gaan dan in op aandachtspunten en aanbevelingen met betrekking tot het voorbereiden en uitvoeren van de toekomstige evaluaties (7.2).

7.1 Evaluatiemethode

Zoals in hoofdstuk 4 is toegelicht biedt een **combinatie van contribution analysis en process tracing** (CA+PT) een robuuste basis voor de evaluatie van de doeltreffendheid en doelmatigheid van NGF-projecten. De gecombineerde aanpak is krachtig in een context waarin [1] er sprake is van parallelle actielijnen en veel (verschillende typen) betrokken partijen, [2] er niet gewerkt kan worden met duidelijke experimentele groepen en controlegroepen op het niveau van het gehele NGF-project, [3] beoogde uitkomsten niet altijd even goed te kwantificeren zijn, en [4] er een Theory of Change ligt die als basis kan dienen voor de evaluatie. Een aanpak die volledig leunt op het bestaan van een counterfactual (op projectniveau) lijkt niet haalbaar in deze context. Alternatieve aanpakken zoals outcome mapping/harvesting zouden in de toekomst weliswaar mogelijk zijn, maar zijn minder toegespitst op het onderzoeken van (beoogde) causale relaties zoals die in een Theory of Change zijn opgesteld. Alle opties meewegende kan de CA+PT-aanpak zowel voor de projecten via de subsidieroute als voor de projecten via de departementale route als basis genomen worden. Voor de projecten in de departementale route moet wel verder verkend worden of er voldoende data en informatie beschikbaar zijn om een dergelijke aanpak goed uit te kunnen voeren.

Met deze aanpak kan een evaluatieontwerp neergezet worden dat in staat is om **causaliteit op een robuuste manier te onderbouwen**. Een veelgebruikte maat om iets te zeggen over de robuustheid van een evaluatieaanpak is de Maryland Scientific Methods Scale (SMS). Hoewel deze SMS doorgaans gebruikt wordt om de robuustheid van (quasi-)experimentele evaluatieontwerpen te beoordelen, zouden we de hier beschreven aanpak op niveau 3 (of mogelijk zelfs 4) op de 5-puntsschaal kunnen

plaatsen. Hoewel de CA+PT-aanpak feitelijk gezien geen counterfactual heeft, biedt de aanpak wel mogelijkheden om systematisch geloofwaardig bewijs op te kunnen bouwen voor de te onderzoeken causale ketens, en kunnen alternatieve verklaringen uitgesloten worden. De robuustheid van de totale aanpak hangt uiteraard af van de kwaliteit van de ontwikkelde testen en de gebruikte bijbehorende data.

Binnen deze aanpak kan een **grote diversiteit van analysetechnieken, dataverzamelingstechnieken en databronnen** ingezet worden. De twee behandelde projecten in hoofdstukken 5 en 6 illustreren deze diversiteit. Het is sterk contextafhankelijk wat mogelijk is en wat de voorkeur geniet. Het is lastig om een algemeen 'keuzemenu' hiervoor op te stellen. Wel is het in de regel zo dat het nuttig kan zijn om testen met behulp van verschillende databronnen uit te voeren om zo de robuustheid van uitkomsten te vergroten ('trianguleren'). Als bijvoorbeeld zowel registerdata, de voortgangsrapportages als interviews dezelfde ontwikkeling onderschrijven kan dit een robuustere basis bieden voor de process tracing test dan dat het geval geweest zou zijn als enkel één van deze bronnen meegenomen zou worden.

Een algemeen aandachtspunt voor de toekomstige NGF-evaluaties is **de timing van de evaluaties in relatie tot de looptijd en fase waarin het NGF-project zich bevindt**. Voor het toetsen van de veronderstelde causale relaties is het immers essentieel dat deze relaties überhaupt de kans hebben gehad om zich te manifesteren. Op het moment dat projecten zich nog in een vroege fase bevinden en er, ook volgens de planning, geen of nauwelijks (intermediate) outcomes gerealiseerd zullen zijn kunnen evaluaties er ook niet in voorzien om deze causaliteit te toetsen. Concreet betekent dit dat tussentijdse evaluaties een focus zullen krijgen op de 'linkerkant' van de Theory of Change: er kan gekeken worden naar de inputs, throughputs en outputs, en mogelijk al naar één of enkele intermediate outcomes. In de evaluatieontwerpen voor Groeien met Groen Staal en Charging Energy Hubs zijn deze beperkingen vanwege timing ook duidelijk te zien. Wel is het mogelijk om vooraf al een goed beeld te krijgen van wat wél mogelijk is in bijvoorbeeld 2026 of 2031. Zo kunnen we nu in 2024 al voorzien wat er grofweg mogelijk zal zijn voor de evaluaties van GGS en CEH in 2026. De partij die een evaluatie wil laten uitvoeren (veelal het ministerie van EZ in deze context) kan hier aan de voorkant al rekening mee houden door een realistische scope te hanteren voor de evaluatie en realistische onderzoeksvragen voor het evaluatieonderzoek te formuleren.

Een ander algemeen aandachtspunt met betrekking tot de dimensie tijd is dat **het realiseren van de uiteindelijke impact (en outcomes) soms pas na de projectlooptijd plaatsvindt**. Het project CEH is hier een voorbeeld van. Deze kanttekening wordt ook door het CPB geplaatst bij dit project³⁹. In dit geval kunnen de projectlooptijden niet aangehouden worden om de volledige (beoogde) impact te evalueren. In deze scenario's liggen twee opties voor de hand, namelijk [1] de evaluatie wel uitvoeren aan het

³⁹ CPB-analyse voorstellen Nationaal Groeifonds (2023), Derde beoordelingsronde, 2023

einde van de projectperiode (en daarbij expliciet extra aandacht besteden aan de plausibiliteit van de pathways richting outcomes en impact), of [2] later in de tijd pas een volledige evaluatie uitvoeren.

Gegeven het resultaat dat een CA+PT-aanpak een goede basis biedt voor de toekomstige evaluaties, zijn er een aantal specifieke aandachtspunten en aanbevelingen die zich richten op de voorbereiding en uitvoering van de evaluaties. Deze lichten we hieronder toe. In 7.3 worden de aanbevelingen samengevat weergegeven.

7.2 Aandachtspunten & aanbevelingen

Het voorbereiden en uitvoeren van complexe programma's en projecten zoals de 50 NGF-projecten komt niet zonder uitdagingen. Ook bij het ontwikkelen van de twee evaluatieontwerpen voor GGS en CEH zijn verschillende aandachtspunten naar boven gekomen. De geïdentificeerde aandachtspunten gelden daarbij niet alleen voor deze twee projecten, maar kunnen ook voor de evaluaties van de NGF-projecten via de departementale route ter harte genomen worden. We hebben deze aandachtspunten gestructureerd aan de hand van twee dimensies: [1] de fase van de evaluatie (voorbereiding/uitvoering) en [2] de onderdelen van het evaluatieontwerp. Hieronder worden de aandachtspunten visueel samengevat. Deze worden in de volgende secties één voor één nader toegelicht.

	 Voorbereiding evaluatie	 Uitvoering evaluatie
1. Theory of Change	• Aanscherpen ToC	• Relevantie kwaliteit ToC
2. Data(verzameling)	• Voorbereiding data <i>binnen</i> kaders project • Voorbereiding data <i>buiten</i> kaders project	• Afhankelijkheid beschikbare data • Denk na over een 'plan B'
3. Analysetechnieken	• Verbeteren PT-testen	• Verbeteren PT-testen • CA/PT-analyse gebaat bij casus-specifiek inzicht
4. Synthese & rapportage	• Scope evaluatie NGF-subsidie	• Omgang met complexiteit en nuances • Aandacht voor 'doorwerk-effect' • Impact NGF ≠ impact gesubsidieerde projecten

Figuur 26 Aandachtspunten m.b.t. de voorbereiding en uitvoering van de evaluaties

7.2.1 Theory of Change

De Theory of Change dient als basis voor het evaluatieontwerp. **Hoe beter de Theory of Change is uitgewerkt en hoe beter de operationalisering ervan middels concrete meetbare indicatoren, hoe beter het mogelijk is om de doeltreffendheid en doelmatigheid van het project goed te kunnen analyseren.** In de twee voorbeelden van GGS en CEH worden de implicaties van een beperkt uitgewerkte Theory of Change (op onderdelen) direct zichtbaar. In het project CEH zijn voor verschillende outcomes en impacts geen duidelijke indicatoren benoemd (zie 6.2.4) en is soms ook niet geheel duidelijk wat er exact verstaan wordt onder een beschreven 'blokje'. Het gevolg is dat evaluatoren zelf een vertaling moeten gaan maken van hetgeen dat beschreven is in het projectplan naar geoperationaliseerde concepten en meetbare indicatoren. Het risico bestaat dat (potentiële) causale relaties niet duidelijk benoemd, beschreven en gemeten worden. In de casus van CEH hebben wij voor het analyseren van een aantal 'blokjes' en 'pijltjes' (PT-tests) weliswaar suggesties gedaan, maar twee belangrijke nadelen hiervan zijn dat [1] dit niet vanuit het consortium zelf gekomen is en [2] er inherent ook geen indicatoren/KPI's geformuleerd zijn waarop het consortium de voortgang gaat meten.

Een oplossingsrichting is om **in de periode voorafgaand aan de evaluatie de Theory of Change aan te scherpen** (waar nodig). Het aanscherpen hiervan dient niet alleen de kwaliteit van de beoogde evaluatie, maar kan ook het consortium helpen in [1] het scherpstellen van wat ze willen bereiken en hoe ze daar kunnen komen en [2] hoe ze dit kunnen monitoren en eventueel bijsturen indien nodig. We bevelen daarom de verschillende departementen en consortia aan om voorafgaand aan de toekomstige (tussentijdse) evaluaties in gesprek te gaan over de Theory of Change met als doel deze verder te concretiseren en aan te scherpen. Eventueel kan hierbij gebruik gemaakt worden van de benadering 'lerend evalueren'.

7.2.2 Data(verzameling)

Om de beschreven Theory of Change goed te kunnen onderzoeken liggen data aan de basis voor het kunnen uitvoeren van de evaluatie. Het niet beschikbaar hebben van de juiste data betekent dat we geen goede uitspraken kunnen doen over gemaakte ontwikkelingen en de effecten die met/door het NGF-project tot stand zijn gekomen.

Een belangrijke bron voor de evaluaties (van de subsidieprojecten) zijn de voortgangsrapportages. In deze voortgangsrapportages wordt aan RVO teruggekoppeld hoe het project is gevorderd op haar inputs, throughputs en outputs (deliverables en milestones). Daarnaast beschrijven de subsidieontvangers diverse KPI's die betrekking hebben op de outcomes en impacts. **Voor het kunnen uitvoeren van de toekomstige evaluaties is het wel van essentieel belang dat de subsidieontvangers ook een duidelijke terugkoppeling geven op alle in het projectplan beschreven indicatoren en KPI's** (zie bijlagen 1 en 2). Om de kans te vergroten dat de benodigde data er in de toekomst

liggen **bevelen we aan om vanuit RVO met de subsidieontvangers om tafel te gaan zitten om voor iedere indicator door te nemen hoe ze deze informatie gaan verzamelen en aan gaan leveren.** Eventuele knelpunten hierbij kunnen dan vroegtijdig geïdentificeerd worden, waardoor er nog de mogelijkheid bestaat om als RVO en subsidieontvanger te kijken naar manieren om de data te kunnen verzamelen en terug te kunnen koppelen in de toekomst (of een alternatief te zoeken). Dit voorkomt dat evaluatoren op het moment van evalueren er tegenaan lopen dat de benodigde data niet beschikbaar zijn en dat er geen mogelijkheden meer bestaan om die data tijdig te verkrijgen. Hierbij geldt de kanttekening dat de voortgangsrapportages hun basis vinden in de subsidieverplichtingen die onder de verleende subsidie liggen. De verplichtingen die subsidieontvangers hebben ten aanzien van het rapporteren beïnvloeden dus in substantiële mate de beschikbaarheid van informatie die benodigd is om een goede evaluatie uit te kunnen voeren.

Een tweede aanbeveling aan de toekomstige evaluatoren is om er rekening mee te houden dat **een 'plan B' voor de dataverzameling van indicatoren ontwikkeld moet worden.** Indien bepaalde informatie bijvoorbeeld niet goed is beschreven in de voortgangsrapportages, zou een plan B ('fallback option') kunnen zijn om middels interviews de benodigde informatie op te halen. Hoewel dit plan B naar verwachting een lagere kwaliteit data oplevert, kan het hebben van enige informatie toch de voorkeur genieten boven helemaal geen informatie hebben. Hoe eerder de evaluatoren zicht hebben op welke data beschikbaar zijn, hoe eerder zij kunnen nadenken over een goed plan B en hoe groter de kans op een robuuste evaluatie. **Idealiter vindt er een 'data assessment' plaats voorafgaand** aan het aanbesteden van het evaluatieonderzoek, opdat toekomstige evaluatoren het evaluatieontwerp hierop kunnen aanpassen. De betrokken departementen en uitvoeringspartijen zouden hier het voortouw in kunnen nemen.

Tot slot zijn er onderdelen van de Theory of Change waar geen indicatoren voor opgesteld zijn. GGS is hier binnen het NGF vermoedelijk een (positieve) uitzondering, omdat zij op alle onderdelen indicatoren hebben benoemd. Bij CEH zijn er al diverse onderdelen zonder duidelijke indicatoren, en onze inschatting is dat de Theory of Change met bijbehorende indicatoren voor projecten in de departementale route over het algemeen minder goed zijn uitgewerkt dan in deze twee subsidieprojecten. Dat heeft er mede mee te maken dat men voor het verkrijgen van de subsidie het project in meer detail uitgewerkt diende te hebben. Om goede evaluaties uit te kunnen gaan voeren, zeker binnen de departementale route, wordt het **nuttig of zelfs nodig om indicatoren te ontwikkelen die niet binnen de projectplannen zijn opgenomen.** Dit kan deels door de consortia zelf opgepakt worden, maar we zien ook mogelijkheden voor uitvoeringsorganisaties zoals RVO en voor de departementen zelf om hier een rol in te spelen. Er zijn immers bepaalde onderwerpen die in veel NGF-projecten terugkomen, zoals kennisontwikkeling, inzet op human capital, en het ontginnen van nieuwe markten. Het is voor deze onderwerpen wellicht mogelijk om **met behulp van generieke databronnen project-specifieke indicatoren te ontwikkelen.** Denk bijvoorbeeld aan DUO-data ten aanzien van instroom van studenten, waar je met deze generieke bron de instroom van

studenten in voor specifieke NGF-projecten relevante opleidingen kunt bepalen. Andere voorbeelden zijn generieke databases met betrekking tot publicaties, patenten, start-ups, en diverse databestanden binnen de CBS-microdata (bijv. Algemeen Bedrijfsregister, Productiestatistiek). Door deze bronnen voor meerdere NGF-projecten in te zetten kunnen **schaalvoordelen** behaald worden. Concreet kan het de benodigde tijd en middelen om nuttige indicatoren te ontwikkelen verkleinen en kan bepaalde benodigde specialistische kennis gecentraliseerd worden (denk bijv. aan kennis over specifieke databronnen of het werken binnen de omgeving van het CBS).

7.2.3 Analysetechnieken

Een centraal punt binnen de analysetechnieken voor de beschreven evaluatieontwerpen zijn de process tracing testen. Hoe sterker de testen, hoe krachtiger het bewijs voor het (niet) bestaan van een causale relatie tussen twee elementen in de Theory of Change. Zoals in hoofdstukken 5 en 6 is benoemd zijn er voor alle impact pathways voorstellen gedaan voor PT-testen, én is het **mogelijk om deze PT-testen aan te scherpen of te 'upgraden'**. Zo biedt bijvoorbeeld een 'Doubly Decisive test' meer bewijskracht dan een 'Hoop-test', dus als het mogelijk is om in nog meer situaties een (praktisch te toetsen) 'Doubly Decisive test' te formuleren vergroot dat de kwaliteit van het evaluatieontwerp.

Een belangrijk aandachtspunt voor het opstellen van een goede CA/PT-analyse is dat **deze analyse gebaat is bij casus-specifieke kennis en inzicht in casus-specifieke bronnen**. Dat vergt domeinkennis van evaluatoren en/of goede voorbereiding vanuit het consortium, het betrokken departement en de betrokken uitvoeringsorganisaties. Hoe meer inzicht in de betrokken casus en de mogelijkheden om zaken te meten, hoe beter het mogelijk is om een robuust ontwerp vorm te geven. Een concreet voorbeeld waarin wij als onderzoekers/evaluatoren in dit traject tegen onze eigen beperkingen aan zijn gelopen is dat we in de casus van Groeien met Groen Staal bijvoorbeeld niet exact weten welke informatie er bij bepaalde inspecties aanwezig is en beschikbaar gesteld zou kunnen worden die nuttig zou kunnen zijn voor het vormgeven van PT-testen. Een goede voorbereiding vanuit de departementen en de consortia met betrekking tot relevante (beschikbare) data kan toekomstige evaluatoren helpen bij het goed kunnen vormgeven en uitvoeren van sterke PT-testen.

7.2.4 Synthese & rapportage

Voor de uiteindelijke synthese & rapportage zijn in dit onderzoek enkele aandachtspunten naar voren gekomen. Eén daarvan heeft betrekking op de uitspraken die je kunt doen over de doeltreffendheid en doelmatigheid van een complex NGF-project. Er zijn doorgaans immers verschillende actielijnen, werkpakketten, en velerlei impact pathways. De kans is groot dat er geen algemene uitspraak gedaan kan worden in de vorm van "het NGF-project is wel/niet doeltreffend". Het kan immers zo zijn dat bepaalde onderdelen van het project wel doeltreffend zijn gebleken en andere

onderdelen niet. Bepaalde impact pathways hebben zich mogelijk wel voltrokken, anderen niet. Het vellen van één algemeen oordeel zou vereisen dat ofwel het NGF-project op *alle* onderdelen doeltreffend geweest is of dat het op *geen enkele* onderdeel doeltreffend is geweest. Die kans achten wij klein. Het ligt dus voor de hand dat er een meer **genuanceerde conclusie getrokken dient te worden, waarbij er oog is voor verschillende onderdelen en impact pathways binnen de Theory of Change**. Een kanttekening die binnen dit onderzoek genoemd is, is dat dit de deur ook open kan zetten voor (ongewenste) ‘politiek gekleurde framing’ en ‘cherry picking’. Dit is in feite voor (vrijwel) iedere evaluatie van toepassing, maar het complexe en veelzijdige karakter van NGF-projecten maakt het inderdaad relatief moeilijker te doorgronden en eenvoudig te beschrijven wat de impact is geweest. We bevelen daarom wel aan om duidelijk te communiceren op welke impact pathways het NGF-project wél doeltreffend is geweest en op welke niet of minder. Ook voor mensen die niet diep in de materie zitten dienen de conclusies begrijpelijk te zijn.

Voor het synthetiseren van de uitkomsten en het rapporteren is het belangrijk om scherp in het vizier te houden **dat voor het doeltreffend kunnen zijn van de NGF-subsidie er op zijn minst sprake moet zijn van input-additionaliteit en/of throughput-additionaliteit**. In andere woorden: de NGF-subsidie moet ervoor gezorgd hebben dat er opgeteld meer input is ingebracht en/of dat er activiteiten zijn ontplooid die zonder de subsidie niet/kleiner/anders/later waren uitgevoerd. Op basis hiervan vindt er immers een doorwerking plaats op de outputs, outcomes en impact. Als men exact hetzelfde had gedaan zonder de subsidie, kunnen eventuele positieve ontwikkelingen op het outcome- en impactniveau niet aan de NGF-subsidie toegekend worden.

Een algemene kanttekening is dat de **huidige opzet van de evaluatie van de NGF-subsidie zich beperkt tot de doeltreffendheid en doelmatigheid van de twee gesubsidieerde projecten**. De NGF-subsidieregeling als zodanig kan in bredere zin effecten teweeggebracht hebben. Het is bijvoorbeeld denkbaar dat niet-gehonoreerde projecten door ‘de kans op subsidie’ wel ontwikkeld zijn en mogelijk later ook (eventueel in lichtere vorm) uitgevoerd zijn. Het bestaan van de subsidie kan consortiumvorming hebben gestimuleerd, en die consortia kunnen ook buiten het NGF om effect sorteren. Er kunnen ook negatieve neveneffecten van de NGF-subsidie bestaan. Partijen waarvoor het voorstel is afgewezen zouden een verlies kunnen ervaren in vertrouwen en/of motivatie in het samen optrekken met de overheid. Ook hadden zij andere, mogelijk productievere, dingen kunnen doen met de geïnvesteerde tijd en middelen die nu in de voorbereiding van het NGF-project zijn gaan zitten. Dit zijn enkele voorbeelden van mogelijke effecten van de subsidieregeling in bredere zin. Voor EZ en andere departementen is het in ieder geval goed om bewust te zijn dat het evalueren van de gehonoreerde NGF-subsidieprojecten een belangrijk, maar niet volledig, beeld geeft van de effecten van de NGF-subsidieregeling.

7.3 Overzicht aanbevelingen

Hieronder worden de belangrijkste aanbevelingen zoals beschreven in 7.2 kort op een rij gezet.

Aanbevelingen

Specifieke aanbevelingen voor Groeien met Groen Staal + Charging Energy Hubs

1. Ga als RVO en consortia in gesprek over de benoemde indicatoren in het projectplan, en maak duidelijke afspraken (bij voorkeur per indicator) over of en hoe deze verzameld gaan worden en hoe hierover gerapporteerd gaat worden in de voortgangsrapportages;
2. Voer als EZ (en RVO) voorafgaand aan de evaluaties een data assessment uit, en houd er als toekomstige evaluatoren rekening mee dat er (op delen van) het hier gepresenteerde voorstel voor een evaluatieontwerp een plan B voor dataverzameling ontwikkeld moet worden;
3. Probeer als toekomstige evaluatoren (en eventueel als EZ en RVO) de gepresenteerde process tracing testen te verbeteren met als doel het verder verstevigen van de robuustheid van het evaluatieontwerp.

Algemene aanbevelingen voor evaluaties van NGF-projecten

4. Scherp als verantwoordelijke departementen en NGF-consortia vroegtijdig de Theory of Change aan gezien de relevantie hiervan voor een robuuste evaluatie;
5. Verken als verantwoordelijke departementen of er op basis van generieke databronnen specifieke indicatoren voor meerdere NGF-projecten ontwikkeld kunnen worden;
6. Streef als toekomstige evaluatoren bij het trekken van conclusies over de doeltreffendheid in de complexe contexten van NGF-projecten naar een goede balans tussen eenvoud en uitlegbaarheid enerzijds en nuance anderzijds;
7. Houd er bij het uitvoeren van de NGF-evaluaties op projectniveau en de algehele evaluatie van het NGF rekening mee dat het evalueren van de gehonoreerde NGF-projecten een niet-volledig beeld geeft van de mogelijke effecten van de NGF-subsidie en het NGF als geheel;

7.4 Slotwoord

Met de voorliggende evaluatiemethode van NGF-projecten is een solide basis gelegd voor toekomstige evaluaties. De CA/PT-aanpak kan in feite voor ieder NGF-project ingezet worden. De uitdaging zal met name zitten in het geven van invulling aan deze aanpak. Belangrijke vragen daarbij zijn “is de Theory of Change voldoende uitgewerkt?”, “zijn er zinvolle indicatoren opgesteld?”, “zijn de benodigde data beschikbaar?” en “kunnen we sterke en praktische haalbare PT-testen formuleren?”. Binnen de subsidieprojecten zijn de projectplannen op relatief veel detailniveau uitgewerkt. Voor de departementale projecten zal het geven van invulling aan de CA/PT-aanpak mogelijk uitdagender zijn, gelet op de complexiteit en omvang van deze projecten. In dit rapport zijn hiervoor reeds enkele aanbevelingen gedaan gelet op toekomstige evaluaties..



Onderzoek voor *onderbouwd* beleid.

Dialogic innovatie & interactie

Hooghiemstraplein 33

3514 AX Utrecht

030-2150580

www.dialogic.nl