

Effectief ontwikkelingsbeleid vergt cultuur van leren en betere evaluaties

Het Nederlandse ministerie van Buitenlandse Zaken spant zich in om de impact van ontwikkelingsprojecten te (laten) evalueren. Momenteel worden monitoring en evaluaties hoofdzakelijk ontworpen voor verantwoordingsdoeleinden, en zijn het soms alleen maar ‘afvinkoefeningen’. Hoe zou dit beter kunnen?

IN HET KORT

- De huidige praktijk van evalueren schiet tekort, onder andere omdat onderzoekers er te laat bij worden betrokken.
- Door meer impact-evaluaties te gaan ontwerpen voor leerdoel-einden kan men tot effectiever ontwikkelingsbeleid komen.
- In een partnerschap kunnen beleidsmakers en onderzoekers samenwerken aan een cultuur van leren.

MARK TREURNIET

Postdoctoraal onderzoeker aan Wageningen University & Research (WUR)

ARJUN BEDI

Hoogleraar aan de Erasmus Universiteit Rotterdam (EUR)

ERWIN BULTE

Hoogleraar aan de WUR

PATRICIO DALTON

Universitair hoofd-docent aan Tilburg University

GESKE DIJKSTRA

Hoogleraar aan de EUR >

De Nederlandse overheid investeert jaarlijks circa 4,5 miljard euro in ontwikkelingssamenwerking om de wereldwijde armoede te bestrijden, en tracht de impact van deze samenwerking te (laten) evalueren. Dit gebeurt deels via een eigen onafhankelijke evaluatiedienst, Internationaal Onderzoek en Beleidsevaluatie (IOB), maar wordt ook uitbesteed aan adviesbureaus, individuele consultants en universiteiten.

Er zijn twee belangrijke redenen om tot monitoren en evalueren over te gaan (IOB, 2009). Ten eerste worden monitoring en evaluaties opgezet voor verantwoordingsdoeleinden, bijvoorbeeld ten behoeve van de bijdrage van Nederland aan het bereiken van de Duurzame Ontwikkelingsdoelen (zie bijvoorbeeld het International Aid Transparency Initiative). In dat geval wordt er gekeken of het geld daadwerkelijk besteed is aan de geplande activiteiten, en hoeveel mensen er daarmee bereikt zijn.

Ten tweede kunnen evaluaties bedoeld zijn om te leren welke interventies er wel werken en welke niet, en hoe dat komt. Hierbij kunnen we onderscheid maken tussen de beleidsevaluaties en impact-evaluaties van concrete projecten. Beleidsevaluaties betreffen sectoren, beleidsthema's of landenprogramma's, en hebben vaak het karakter van een synthesestudie. Bij deze beleidsevaluaties spelen literatuur en de beschikbare impact-evaluaties een belangrijke rol (IOB, 2009). Helaas blijken de beschikbare impact-evaluaties hiervoor vaak onvoldoende geschikt (IOB,

2019). Deze evaluaties zijn soms een verplicht nummer, iets wat nu eenmaal moet gebeuren, en worden niet georganiseerd om het leren te maximaliseren. Daardoor blijven de kansen om te leren onbenut.

Om leren mogelijk te maken, zijn betere impact-evaluaties vereist, met een andere evaluatiestrategie – een die gericht is op het verkrijgen van inzicht in causale verbanden en onderliggende mechanismen. Esther Duflo, Nobelprijswinnaar in 2019, benadrukt het belang hiervan in haar acceptatielezing (Duflo, 2020), en ook in het *ESB*-thema-nummer 'Impact in het mondiale Zuiden' (2020) worden er een aantal mooie voorbeelden van effectiviteitsmetingen gepresenteerd.

Recente discussies over ontwikkelingsbeleid schenken dan ook veel aandacht aan het belang van op feiten gebaseerde beslissingen door beleidsmakers. Om tot een effectiever beleid te komen, zijn buitenlandse diensten zoals de *Norwegian Agency for Development Cooperation* (Norad), het Britse *Foreign, Commonwealth & Development Office* (FCDO) en de Wereldbank steeds meer gaan inzetten op het leren via evaluaties (Wereldbank, 2021).

Waar het Nederlandse Ministerie van Buitenlandse Zaken een wegbereider was door het (laten) uitvoeren van impact-evaluaties om van te leren, speelt ze die voorttrekkersrol nu al lang niet meer. Het ministerie erkent dit en is bezig om de capaciteit voor monitoring, evaluatie en leren te vergroten, zodat leeragenda's een meer centrale rol kunnen spelen in de samenwerking met partners (BZ, 2019). Bij deze bijsturing wordt er echter onvoldoende ingezet op impact-evaluaties, waardoor leren op de lange termijn beperkt blijft.

Wij pleiten daarom voor meer ruimte om leerdoel-einden systematisch te evalueren. Via beter opgezette evaluaties, gecombineerd met een sterke cultuur van leren en samenwerken, kan de effectiviteit van het Nederlandse ontwikkelingsbeleid worden verbeterd.

Criteria bij grote leerbatens

Er zijn drie criteria die helpen bepalen waar de 'leerbatens' van een goede effectmeting het grootst zijn (Gertler et al., 2016). Het eerste criterium is dat de ontwikkelings-interventie nog niet goed onderzocht is, zodat er weinig bekend is over de effectiviteit van het programma (zie bijvoorbeeld de *Evidence Gap Maps* (3ie, 2021)), en hoe de effectiviteit afhangt van de implementatie en de specifieke context.

Het tweede criterium betreft de kennisrelevantie: de evaluatie bouwt voort op bestaande theoretische en empirische kennis, en helpt zo om te begrijpen waarom de interventie wel of niet een impact realiseert.

Het derde criterium betreft de potentiële invloed van de evaluatie: de resultaten worden gebruikt wanneer er belangrijke beleidsbesluiten worden genomen. Dit vereist van betrokken partijen partnerschap en commitment vanaf het begin, om zo open te staan voor daadwerkelijk leren, en ook een tijdige beschikbaarheid van de resultaten (Pattyn en Bouterse, 2020).

Recente ervaringen met een evaluatie van een door het Ministerie van Buitenlandse Zaken gefinancierd publiek-privaat partnerschap, waarbij twee van ons betrokken waren, illustreren dat er niet altijd wordt voldaan aan de voorwaarden qua leerbatens (kader 1).

Het directe gevolg hiervan is dat deze evaluatie weinig heeft bijgedragen aan de leerdoelstellingen. Hoe zouden we nu betere antwoorden kunnen krijgen?

Investeer in degelijke methodologie

De vraag hoe effectief een interventie is, is makkelijker gesteld dan beantwoord. Hiervoor moet de situatie van de groep die een interventie ontvangt vergeleken worden met wat er gebeurd zou zijn als deze groep de interventie niet zou hebben gekregen. Deze zogenaamde *counterfactual* kunnen we niet observeren, en dat is precies wat impact-evaluaties zo uitdagend maakt.

Schatting van de counterfactual

Om de counterfactual te schatten maken impact-evaluaties daarom gebruik van controlegroepen. Vergelijkingen tussen interventie- en controlegroepen worden doorgaans echter geplaagd door *selectie-effecten*: mensen die deelnemen aan een interventie zijn meestal anders dan mensen die dat niet doen. Als we de verschillen in uitkomsten observeren tussen de interventie- en de controlegroep, dan is het dus de vraag of dit het gevolg is van de interventie, of dat de twee groepen bij aanvang al anders waren. Een degelijk ontwerp voor de impact-evaluatie kan dit probleem oplossen.

Wanneer de groep mensen die in aanmerking komt voor een programma groter is dan het aantal (initieel) beschikbare plaatsen, dan kan er *willekeurige toewijzing* gebruikt worden: een loterij zou kunnen bepalen wie de interventie (in eerste instantie) krijgt. Behalve dat dit mensen een gelijke kans geeft op deelname, zorgt dit ervoor dat er twee groepen ontstaan die vergelijkbaar zijn op een enkel verschil na: één groep ontvangt de interventie, en de andere groep (nog) niet.

Als willekeurige toewijzing niet mogelijk is (Deaton, 2020), dan kan er op basis van kennis over de interventie-implementatie in sommige gevallen worden gewerkt met een selecte, niet-gerandomiseerde controlegroep. In dat geval zullen we echter op zijn minst willen weten of de eventuele verschillen die we aantreffen bij de eindmeting niet ook al bestonden voordat de interventie plaatsvond. Het verzamelen van *baseline*-data is dan noodzakelijk.

Het tijdig verzamelen van baseline-data voor de juiste groepen is echter lastig als de implementatieplannen relatief lang onzeker blijven. Bij de grootschalige evaluatie van het

Nederlandse Medefinancieringsprogramma MFS-II werden onderzoekers bijvoorbeeld pas betrokken toen de meeste interventies al gestart waren, waardoor willekeurige toewijzing geen optie meer was, en er geen baseline-data meer konden worden verzameld (Van der Gaag et al., 2015).

Evaluaties zouden dus een stuk betrouwbaarder en waardevoller worden als onderzoekers vanaf het allereerste werden betrokken bij de gesprekken, om zo mee te kunnen denken over de implementatie van de interventie en het evaluatiedesign.

Voldoende steekproefomvang

Zolang evaluatiebudgetten over veel kleine projecten worden verdeeld, is er meestal te weinig geld voor voldoende grote steekproeven om de eventuele effecten te kunnen vaststellen. Van der Gaag et al. (2015) suggereren dan ook dat als budgetten beperkt zijn, men beter kan kiezen voor minder evaluaties dan voor te kleine steekproeven. Dit onderstreept de noodzaak om in te zetten op de belangrijkste leervragen.

Complementaire methoden

In het algemeen is een combinatie van kwalitatieve methoden, beschrijvende statistiek en kwantitatieve impactschattingen essentieel om samenhangende leervragen te beantwoorden: diverse methoden hebben elk hun eigen sterke en zwakke punten, en de verschillende sterktes kunnen in een zogeheten *mixed methods*-design gecombineerd worden.

Garandeer toegang tot administratieve data

Uitkomsten worden gemeten met behulp van primaire data (denk aan huishoud-enquêtes), maar vaak zijn aanvullende secundaire data (denk aan administratieve data) ook zeer nuttig – en soms zelfs onmisbaar. Uitvoerende partners hebben echter niet altijd direct baat bij onafhankelijk onderzoek, en zien soms – met een beroep op ‘strategische redenen’ of ‘privacyoverwegingen’ – mogelijkheden om te voorkomen dat bepaalde onderzoeksvragen beantwoord worden door de benodigde data niet te delen. In het in kader 1 beschreven geval van het ‘publiek-privaat partnerschap’ konden we de verdelingsvraag slechts beantwoorden dankzij de medewerking van mensen bij niet minder dan vijf verschillende organisaties.

In het geval van interventies die betaald zijn met publiek geld zou de financier van deze programma’s een belangrijke rol moeten spelen bij het maken van afspraken met uitvoerende partners over toegang tot deze data. Dit zou tevens bijdragen aan het Duurzame Ontwikkelingsdoel 17 (‘Versterken van het mondiaal partnerschap om doelen te bereiken’; SDG (2021)).

Organiseer transparantie

Onderzoekskwaliteit is natuurlijk primair een verantwoordelijkheid van de onderzoekers. We noemen echter twee maatregelen die kunnen helpen om de kwaliteit te borgen.

Ten eerste een *peer review* van het evaluatieplan en de uitkomsten: onafhankelijke deskundigen geven feedback om de kwaliteit van onderzoek te borgen en te verbeteren. En ten tweede *replicatie*: andere onderzoekers kunnen de analyse herhalen en controleren op eventuele fouten.

JAN WILLEM GUNNING

Emeritus Hoogleraar aan de Vrije Universiteit Amsterdam (VU)

WENDY JANSSENS

Hoogleraar aan de VU

LUUK VAN KEMPEN

Universitair docent aan de Radboud Universiteit

ROB KUIJPERS

Onderzoeker bij het KIT Royal Tropical Institute

ROBERT LENSINK

Hoogleraar aan de Rijksuniversiteit Groningen

KARLIJN MORSINK

Universitair docent aan de Universiteit Utrecht

ELEONORA NILLESEN

Hoogleraar aan de Universiteit Maastricht

REMCO OOSTENDORP

Hoogleraar aan de VU

MENNO PRADHAN

Hoogleraar aan de VU en de Universiteit van Amsterdam

RUERD RUBEN

Emeritus hoogleraar aan de WUR

DAAN VAN SOEST

Hoogleraar aan Tilburg University

Casus: evaluatie van landbouwproject

KADER 1

De evaluatie van een landbouwproject in een ontwikkelingsland laat zien hoe de (grootste) leerbaten nu niet altijd worden gerealiseerd. De bedoeling van het partnerschap was om leden van een lokale boerencoöperatie te helpen om de kwaliteit van hun productie te verbeteren, zodat zij van een hogere prijs zouden kunnen profiteren. De interventies van het partnerschap hebben inderdaad tot een hogere kwaliteit geleid, maar of boeren daadwerkelijk profiteren van een hogere prijs voor hun verbeterde productie hangt af van de marktmacht van de afnemer. En juist deze vervolgvraag werd niet meegenomen in het originele evaluatieplan, namelijk: hoe wordt de verhoogde toegevoegde waarde van de productie verdeeld tussen marktpartijen?

Via extern gefinancierd vervolgonderzoek hebben we ten eerste de bevinding vast kunnen stellen dat de boerencoöperatie op de korte termijn een hogere prijs kreeg vanwege de verbeterde kwaliteit, maar ook dat dit voordeel relatief snel weer verdween. De baten van de (met publiek geld gefinancierde) interventie kwamen voor het groot-

ste deel terecht bij de (private) afnemer, een internationaal opererend bedrijf.

De tweede bevinding is dat het project bestond uit een combinatie van trainingen, gesubsidieerde inputs en het invoeren van individuele bonussen voor betere kwaliteit – die tezamen op de voedselkwaliteit een positieve impact hadden. Omdat alle boeren dezelfde combinatie van deelinterventies kregen, was het niet mogelijk om de afzonderlijke bijdragen van trainingen, subsidies en bonussen te bepalen. Daardoor weten we eigenlijk nog steeds niet in welke combinatie de verschillende deelinterventies optimaal ingezet en eventueel opgeschaald kunnen worden.

De derde bevinding is dat de beleidsmaker alleen over de resultaten wilde praten als ook de betrokken private partner zou aanschuiven. De private partner zag echter geen aanleiding om in gesprek te gaan, en dus was er van gezamenlijk leren geen sprake. Onze resultaten worden weliswaar gepubliceerd in de internationale vakliteratuur, maar blijven in beleidskringen vooralsnog onbesproken.

Voor peer review en replicatie is het noodzakelijk om de gebruikte (geanonimiseerde) data te kunnen delen. In het verlengde van het geven van beperkte toegang tot secundaire data staan echter (sommige) uitvoerende partners niet toe om de (geanonimiseerde) data te delen voor replicatiedoelinden. Wederom zou de financier een faciliterende rol kunnen spelen door hier vooraf duidelijke afspraken over te maken. Een goed voorbeeld hiervan is de *Millennium Challenge Corporation* (MCC, 2021), een Amerikaanse ontwikkelingsdienst naast USAID, die zowel evaluatieresultaten als de onderliggende data publiceert op zijn website.

Borg materiële onafhankelijkheid

Voor de geloofwaardigheid van evaluaties dienen de onderzoekers geen belang te hebben bij de uitkomsten. Evaluaties worden meestal uitgevoerd door 'formeel onafhankelijke' organisaties, maar dat is niet helemaal hetzelfde als 'zonder belang'. Een 'onafhankelijk onderzoeker' die voor vervolgoopdrachten afhankelijk is van de uitvoerder is allicht meer geneigd om mee te denken over onderzoeksrichtingen waarvan er eventueel impact te verwachten is, en over de wijze waarop resultaten gerapporteerd worden.

Bij meerdere projecten zien we dat, als de resultaten tegenvallen, dit voor implementerende partijen een aanleiding is om in gesprek te gaan over de gebruikte methodologie en om het gebrek aan impact ten eerste aldaar te zoeken: 'de evaluatie kwam te vroeg', 'er zijn verkeerde uitkomstvariabelen gebruikt', of 'de controlegroep heeft indirect ook van de interventie geprofiteerd'. Het is dan ook belangrijk om vooraf het evaluatiedesign grondig te bespreken en te accorderen.

Momenteel hebben implementerende partijen vaak een stem bij de keuze van de organisatie die hun inter-

ventie gaat evalueren, en zijn sommige opdrachtnemers afhankelijk van de toekomstige opdrachten van dezelfde partij. Selectieve rapportage komt dus vermoedelijk vaak voor, met alle gevolgen daarvan voor de validiteit van de impact-evaluatierapporten die terechtkomen op de bureaus van beleidsmakers. Een oplossing ligt voor de hand: om de druk op opdrachtnemers te verminderen, zouden ze moeten worden geselecteerd door een onafhankelijke organisatie die expertise heeft op evaluatiegebied.

Creëer een cultuur van leren

Natuurlijk wordt er binnen beleidskringen ook belang gehecht aan leren. Dit wordt bijvoorbeeld georganiseerd via zogenaamde kennisplatforms, en (in het verleden) ook middels onderzoeksprogramma's die worden uitgezet via NWO. Maar wij hebben de indruk dat belangrijke kansen om samen te leren nog onbenut blijven, met name op het gebied van projectevaluaties. Dergelijke evaluaties zijn onvoldoende doordacht, en daarom meestal ongeschikt om beleidsevaluaties te ondersteunen. Intensievere samenwerking tussen beleidsmakers, uitvoerende organisaties, potentiële begunstigen en onderzoekers kan helpen om meer te leren uit evaluaties en die kennis op nieuw beleid toe te passen.

Co-creatie

Over het algemeen is leren van interventies het meest effectief als beleidsmakers en uitvoerende organisaties langdurig samenwerken met potentiële begunstigen en onderzoekers (Dufflo, 2020). Potentiële begunstigen kunnen delen wat ze zelf als oplossingen zien, en hoe hun eigen inspanningen en interventies elkaar zouden kunnen versterken, terwijl de beleidsmakers en uitvoerende organisaties kunnen leren van de effecten die deze projecten hebben. De verschillende partijen kunnen voorstellen doen om met innovaties te experimenteren. En samen kunnen ze interventies, innovaties en onderzoeken zo ontwerpen dat deze zo veel mogelijk helpen om het beleid effectiever te maken.

Bovendien vinden er door samenwerking allerlei vormen van kennisoverdracht plaats. De onderzoekers kunnen aandragen wat al bekend is in de wetenschappelijke literatuur, en hoe daarop kan worden voortgebouwd. Vervolgens kunnen de onderzoekers van de andere betrokkenen leren over de beleidsrelevante context, de (financiële) haalbaarheid van interventies en politieke afwegingen. Zo kan er in dit proces een gezamenlijke cultuur van leren ontstaan, waarin elke partij haar eigen rol en verantwoordelijkheid behoudt, maar waarin innovatieve ideeën getest worden en falen vooral wordt gezien als een kans om te leren het de volgende keer beter te doen.

Experimenteren met innovatie

De leercultuur kan verder worden versterkt door een deel van de publieke financiering te reserveren voor projecten die voldoen aan de eerder besproken criteria van leren, en om een ander (groter) deel van het budget te reserveren en dit te investeren op basis van de nieuw verkregen inzichten in wat er daadwerkelijk effectief blijkt te zijn. Mooie voorbeelden hiervan zijn de faciliteit *Development Impact Ventures* (DIV) van USAID en het *Fund for Innovation in Develop-*

ment (FID) van de Franse ontwikkelingsdienst (USAID, 2021; FID, 2021). Iedereen met een innovatief idee voor armoedebestrijding (zoals bijvoorbeeld een nieuwe technologie, een nieuw businessmodel, of een nieuw beleidsidee) kan bij de DIV en FID een aanvraag doen voor de financiering van een pilot. Als de pilot succesvol is, kan er vervolgfianciering worden aangevraagd om de interventie te testen, en te onderzoeken hoe de interventie kan worden opgeschaald. Vervolgens kan de financiering worden aangevraagd voor de feitelijke uitrol van de meest effectieve, schaalbare interventies.

Partnerschap

Omdat een cultuur van leren niet zomaar ontstaat, pleiten wij voor een intensivering van de relatie door middel van een partnerschap tussen enerzijds beleidsmakers van het Ministerie van Buitenlandse Zaken en anderzijds wetenschappers van verschillende universiteiten. Dat partnerschap zou het model kunnen volgen van het BIG'R-partnerschap tussen de gemeente Rotterdam en de Erasmus Universiteit – een team bestaande uit ambtenaren en wetenschappers ontmoet elkaar op regelmatige basis, wisselt kennis uit, formuleert oplossingen voor concrete problemen op basis van wetenschappelijke inzichten, en test in de praktijk of deze oplossingen daadwerkelijk tot de beoogde resultaten leiden. Een vergelijkbaar initiatief, op nationaal niveau, is het zogenaamde *Behavioural Insights Network Nederland*, waarbinnen beleidsmakers en onderzoekers elkaar ontmoeten.

Naast het opzetten van een partnerschap tussen beleidsmakers en wetenschappers in Nederland is het voor ontwikkelingssamenwerking van groot belang dat ook beleidsmakers, potentiële begunstigden en wetenschappers uit ontwikkelingslanden worden betrokken bij een dergelijk overleg. In gebieden waar veel Nederlandse hulp besteed wordt, zouden observatoria voor onderzoek en leren kunnen worden opgezet, waarin Nederlandse onderzoekers en lokale kennisinstellingen samenwerken met als doel om de effectiviteit van het gevoerde beleid te verhogen. Recentelijk heeft de stuurgroep van WOTRO een pleidooi gehouden voor het opzetten van een dergelijk onderzoeksprogramma in het document *Turning the tide* (WOTRO Stuurgroep, 2020).

Verlammende politieke cultuur

Waarom gebeurt dit allemaal niet? In een discussie met vertegenwoordigers van Buitenlandse Zaken en IOB, geeft men aan een (groot) aantal van de geschetste tekortkomingen te herkennen, en ook dat er recent initiatieven zijn opgestart om meer te leren. Omdat de politiek vaak op korte termijn verantwoording vraagt voor allerlei initiatieven, worden evaluaties al gauw tot een haastklus. Het ministerie verwelkomt het idee om een selectie van (pilot)projecten grondig te evalueren met het doel om te leren, hoewel men tegelijkertijd vreest dat eventuele tegenvallende evaluatie-resultaten (die juist van grote waarde zijn wanneer men de effectiviteit van beleid wil verbeteren) zullen leiden tot politiek gedoe en druk op toekomstige budgetten. De politieke cultuur, met haar nadruk op opportunisme en snelle succesjes, heeft op deze manier een verlamdend effect op

het vermogen van ambtenaren en wetenschappers om te komen tot een effectiever ontwikkelingsbeleid.

Tot slot

Wereldwijde armoede is een complex probleem, en dus is het van het grootste belang om te leren welke oplossingen en innovaties er effectief zijn. Binnen de huidige kaders blijkt het echter niet eenvoudig om de impact van interventies te evalueren, en daar al doende van te leren.

Het blijven uitvoeren van ineffectief beleid is kostbaar, maar degelijke evaluaties zijn dat ook. Daarom is het verstandig om niet zomaar alle of willekeurig geselecteerde projecten te evalueren, maar om beleidsexperimenten en evaluaties specifiek in te zetten waar de leerbaten het grootst zijn. In deze gevallen kan er vervolgens voor worden gezorgd dat er wél voldaan wordt aan belangrijke voorwaarden voor evaluaties – degelijke methodologie, toegang tot data, transparantie en materiële onafhankelijkheid.

Intensivering van de samenwerking tussen beleidsmakers en onderzoekers, zowel in Nederland als in landen waar de meeste Nederlandse interventies plaatsvinden, kan helpen om de hiervoor benodigde cultuur van leren te creëren. Ontwikkelingseconomen van verschillende universiteiten zijn van harte bereid om (samen met experts op het gebied van bijvoorbeeld landbouw, voeding, gezondheid en onderwijs) hieraan bij te dragen – dus aan een partnerschap voor een effectiever ontwikkelingsbeleid.

Literatuur

- 3ie (2021) *Evidence gap maps*. Internal Initiative for Impact Evaluation. Informatie te vinden op www.3ieimpact.org.
- BZ (2019) *Beleidsreactie op evaluatie Wederopbouw tender, Strategische partners chronische crises programma en ARC tender*. Ministerie van Buitenlandse Zaken. Kamerbrief, 32605 nr. 218, 19 oktober. Te vinden op www.tweedekamer.nl.
- Deaton, A. (2020) Introduction: randomization in the tropics revisited, a theme and eleven variations. In: F. Bédécarrats, I. Guérin en F. Roubaud (red.), *Randomized control trials in the field of development: a critical perspective*. Oxford: Oxford University Press, p. 29–46.
- Duflo, E. (2020) Field experiments and the practice of policy. *The American Economic Review*, 110(7), 1952–1973.
- ESB (2020) Impact in het mondiale Zuiden. ESB, 105(4782).
- FID (2021) *Fund for innovation in development*. Informatie te vinden op fundinnovation.dev.
- Gaag, J. van der, J.W. Gunning en G. Rongen (2015) *MFS II joint evaluations final synthesis report*. NWO. Te vinden op www.partos.nl.
- Gertler, P.J., S. Martinez, P. Premand et al. (2016) *Impact evaluation in practice*. Tweede editie. Washington, DC: Inter-American Development Bank / The World Bank.
- IOB (2009) *Evaluatiebeleid en richtlijnen voor evaluaties*. Rapport Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie, oktober. Te vinden op www.iob-evaluatie.nl.
- IOB (2019) *IOB evaluatie: less pretension, more realism*. Rapport Internationaal Onderzoek en Beleidsevaluatie, juli. Te vinden op www.rijksoverheid.nl.
- MCC (2021) *Evaluation*. Millennium Challenge Corporation. Informatie te vinden op www.mcc.gov.
- Pattyn, V. en M. Bouterse (2020) Explaining use and non-use of policy evaluations in a mature evaluation setting. *Humanities & Social Sciences Communications*, 7, artikelnr 85.
- SDG (2021) *Sustainable Development Goals: werelddoelen voor duurzame ontwikkeling*. Informatie te vinden op www.rijksoverheid.nl.
- USAID (2021) *Development innovation ventures*. Informatie te vinden op www.usaid.gov.
- Wereldbank (2021) *DIME Overview*. Informatie te vinden op www.worldbank.org.
- WOTRO Stuurgroep (2020) *Turning the tide*. Te vinden op www.nwo.nl.