

Welke factoren bevorderen of belemmeren het gebruik van beleidsevaluaties?

Resultaten van een studie bij de Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie

Marjolein Bouterse en Valérie Pattyn *

Aanbevolen citeerwijze bij dit artikel

Marjolein Bouterse en Valérie Pattyn, 'Welke factoren bevorderen of belemmeren het gebruik van beleidsevaluaties?', *Beleidsonderzoek Online* maart 2018, DOI: 10.5553/BO/221335502018000003001

1 Inleiding

Ex post of ex durante evaluatieonderzoek dient vaak een dubbel doel: verantwoording afleggen over het gevoerde beleid en het verbeteren van bestaand beleid. Het uitvoeren van een evaluatie is echter geen garantie dat beleidsverbetering ook daadwerkelijk wordt bereikt. Immers, er zijn tal van factoren die beïnvloeden of een evaluatie al dan niet instrumenteel wordt gebruikt voor beleidsaanpassingen. Om deze reden is het gebruik van evaluaties bij uitstek het onderwerp van onderzoekers over evaluaties (Alkin, 2004). Ook in dit tijdschrift werd al herhaaldelijk aandacht besteed aan het gebruik van evaluaties (Korsten, 2013; Korsten & Van der Meer, 2014; Van Hoesel, 2017). Dat evaluaties niet worden opgepikt, kan problematisch zijn, niet in het minst omwille van publieke middelen die dan mogelijk niet optimaal worden aangewend. De kernvraag is bijgevolg te achterhalen *onder welke (combinaties van) voorwaarden gebruik van evaluaties wordt bevorderd of belemmerd*. Dit is de vraag die in dit artikel centraal staat. We focussen hierbij op ex post en ex durante evaluaties; en op instrumenteel gebruik (zie verder).

Ondanks tal van beschikbare waardevolle inzichten over evaluatiegebruik, kent bestaand onderzoek toch nog belangrijke tekortkomingen. Vooreerst is het zo dat veel onderzoek berust op anekdotisch bewijs, wat eigen is aan een praktijkveld als evaluatie (Henry & Mark, 2003). Dergelijk anekdotisch bewijs dat gebaseerd is op ervaringen met een of meerdere evaluaties is uiteraard zeer interessant, maar is problematisch vanuit comparatief perspectief, wanneer men systematische lessen wil trekken over evaluaties heen. Waar comparatief onderzoek wel beschikbaar is (bijv. Barrios, 1986; Bober & Bartlett, 2004; Azzam & Levine, 2015), stellen we vast dat er relatief weinig aandacht is voor de samenhang tussen verschillende factoren die impact kunnen hebben op het gebruik van evaluaties. Uit de meta-analyse van Johnson et al. (2009) blijkt dat 16 van de 41 studies maar één onderzochte factor bevatten. Studies identificeren wel tal van factoren die het gebruik van evaluaties bevorderen, maar of er een welbepaalde samenhang is tussen deze factoren, blijft onduidelijk. Ten derde richt onderzoek zich overwegend op factoren die gebruik bevorderen, en blijven de factoren die gebruik verhinderen vaak uit het zicht. In voorliggend artikel presenteren we de resultaten van een onderzoek waarin we hebben getracht deze tekortkomingen aan te pakken. We hebben hiertoe gebruikgemaakt van Qualitative Comparative Analysis (QCA), een methode die algemeen in opmars is in de evaluatiewereld als instrument voor evaluatieonderzoek (e.g. Pattyn et al., 2017) én voor onderzoek over evaluaties (bijv. Balthasar, 2006; Ledermann, 2012). QCA is uitermate geschikt om het gebruik van meerdere evaluaties systematisch te vergelijken, en de combinaties van factoren

te identificeren die gebruik kunnen bevorderen of verhinderen. Ons onderzoek dient hierbij niet enkel een fundamenteel nut, maar heeft als voornaamste doelstelling lessen aan te reiken om het gebruik van evaluaties zelf te bevorderen.

Het empirisch onderzoek¹ werd uitgevoerd bij het ministerie van Buitenlandse Zaken in Nederland, en richtte zich op de evaluaties die werden uitgevoerd door de Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (IOB)² in de periode 2013-2016. IOB vormt een interessante casus voor het onderzoek naar gebruik van evaluaties, omwille van de lange traditie van evaluatieonderzoek in de ontwikkelingssamenwerking in het algemeen, en de reputatie van IOB als voortrekker op evaluatiegebied in Nederland in het bijzonder. IOB heeft een zeer ruime ervaring ter zake, ook als verkenner van nieuwe evaluatiemethoden. Bovendien heeft IOB het 'leren van evaluaties' voor de eigen organisatie expliciet als doel vooropgesteld (Ministerie van Buitenlandse Zaken, 2009).

Wat betreft de focus op het gebruik van evaluaties, dienen de volgende elementen in het achterhoofd te worden gehouden bij het lezen van dit artikel. In het onderzoek richtten we ons op het gebruik in de vorm van concrete beleidsaanpassingen naar aanleiding van de resultaten van evaluatieonderzoek. Hierbij brachten we enkel de beleidswijzigingen in kaart die werden gemaakt door beleidsmedewerkers op het ministerie van Buitenlandse Zaken; aanpassingen door bijvoorbeeld uitvoeringsinstanties of ambassades werden niet meegenomen. We erkennen uiteraard dat er nog tal van andere vormen zijn van evaluatiegebruik, zoals meer inzicht vergaren (*enlightenment use*, Weiss, 1979) of conceptueel gebruik. Deze vormen werden echter niet bestudeerd. Tot slot willen we benadrukken dat het onderzoek niet zonder meer kan worden gegeneraliseerd naar andere contexten. Dit doet evenwel geen afbreuk aan de relevantie. De studie wijst vooreerst op de noodzaak én de mogelijkheid om factoren systematisch te onderzoeken over diverse evaluaties heen. Het geeft daarenboven interessante aanwijzingen aan evaluatoren en beleidsmakers om in overweging te nemen bij het opzetten en uitvoeren van de evaluatie- en beleidspraktijk. Verder biedt het inzicht in het opzetten van een meer systematische onderzoeksagenda rond het thema.

Dit artikel zal in de volgende sectie ingaan op de gemaakte keuzes in de conceptualisering en op de gebruikte methodologie. In de derde sectie wordt de casus uitgewerkt en worden de verklarende factoren en de afhankelijke variabele geoperationaliseerd. Vervolgens worden in sectie 4 de bevindingen van de QCA beschreven. We sluiten het artikel af met een discussie van de onderzoeksresultaten en de implicaties voor evaluatietheorie en -praktijk.

2 Opzet van het onderzoek

In dit onderzoek wordt de veelgebruikte definitie van het Development Assistance Committee (DAC) van de Organisatie voor Economische Samenwerking en Ontwikkeling (OESO) aangehouden om een evaluatie te definiëren. Dit is ook de definitie die het ministerie van Buitenlandse Zaken hanteert, en was aldus herkenbaar voor de respondenten die in het onderzoek werden betrokken. Volgens deze definitie is een evaluatie: 'een beoordeling, zo systematisch en objectief mogelijk, van het design, de implementatie en resultaten van een lopend of afgerond project, programma of beleid. Hoofddoel is het bepalen van de relevantie van het project en het doelbereik, de ontwikkelingsefficiëntie, effectiviteit, impact en duurzaamheid. Een evaluatie dient informatie op te leveren die betrouwbaar en nuttig is, en laat toe om geleerde lessen op te nemen in de besluitvorming van donoren en ontvangers van ontwikkelingsgelden' (eigen vertaling, Development Assistance Committee, 1991). Alhoewel gelinkt, is deze definitie dus ruimer dan de definitie die gehanteerd wordt in de Regeling Periodiek Evaluatieonderzoek, waar de klemtoon vooral ligt op de evaluatie van doelmatigheid en doeltreffendheid van beleid.

2.1 Conceptuele keuzes inzake evaluatiegebruik en -gebruiker

2.1.1 Instrumenteel gebruik

Wanneer beleidskeuzes expliciet worden gebaseerd op (een) concrete beleidsevaluatie(s), kan men spreken van het instrumenteel gebruik van deze evaluatie(s) (Alkin & Taut, 2003;

Ledermann, 2012). We maken hierbij wel de volgende kanttekeningen. Ten eerste, instrumenteel gebruik volgens onze operationalisering veronderstelt niet per se *beleidsverbetering*. Ten tweede hoeven er evenmin concrete *beleidsveranderingen* te hebben plaatsgevonden. Ook de keuze om de status quo voort te zetten kan worden beschouwd als een beleidskeuze (Dye, 1972) en kan zijn gebaseerd op een beleidsevaluatie (Alkin & Taut, 2003: 5). Het spreekt verder voor zich dat de evaluator het niet per se eens hoeft te zijn met de op de evaluatie gebaseerde keuze, maar dat deze in handen ligt van de beleidsmakers.

2.1.2 Categorieën van gebruik

Om instrumenteel gebruik verder te kaderen baseren we ons overwegend op drie dimensies die Kirkhart (2000) onderscheidt (zie tabel 1). Kirkhart maakt een onderscheid tussen het moment van gebruik vanuit het perspectief van de evaluatie, de bedoeling van gebruik en de zogenaamde bron van gebruik.

Tabel 1 Categorisering van gebruik

Moment van gebruik	Bedoeling van gebruik	Bron van gebruik
Direct	Bedoeld	Proces
Aan het einde van de evaluatie	Onbedoeld & voorzien ^a	Bevindingen
Lange termijn	Onbedoeld & onvoorzien	

^a Kirkhart (2000) onderscheidt enkel 'bedoeld' en 'onbedoeld'. Alkin en Taut (2003) tonen dat onbedoeld verder gespecificeerd kan worden in 'onbedoeld en voorzien' en 'onbedoeld en onvoorzien'.

Wat betreft het *moment van gebruik* kunnen we drie mogelijkheden onderscheiden. Beleidsmakers kunnen ofwel de evaluatie direct gebruiken, i.e. nog terwijl het proces van evaluatie gaande is; zodra de evaluatie klaar is; of op langere termijn. De *bedoeling van gebruik* onderscheidt evaluaties die gebruikt zijn zoals de evaluatoren dat bedoeld hadden, van evaluaties die op een andere wijze zijn gebruikt. De dimensie *bron van gebruik* betreft in feite het object van gebruik: werden vooral de bevindingen van de evaluaties gebruikt of komt het gebruik eerder voort uit inzichten gehaald uit de betrokkenheid bij het evaluatieproces? Betrokkenheid bij het evaluatieproces kan er bijvoorbeeld toe leiden dat beleidsmakers expliciet reflecteren op het beleid of er onderling over gaan spreken. Dergelijke reflectie of discussie kan vervolgens bepaalde beleidskeuzes beïnvloeden, ongeacht de uiteindelijke bevindingen die de evaluatie levert. In dit onderzoek is gekeken naar gebruik dat direct, dan wel aan het einde van de evaluatie heeft plaatsgevonden. Langetermijngebruik kon, omwille van praktische overwegingen, niet worden meegenomen. Verder is er alleen gekeken naar bedoeld gebruik, conform de hierboven beschreven DAC-definitie dat evaluaties instrumenteel gebruik als hoofddoel moeten hebben. Ten slotte is alleen gebruik op basis van de bevindingen meegenomen.

2.1.3 De gebruiker

Zoals we hierboven al suggereerden, verwijzen we in dit onderzoek met 'gebruiker' of 'mogelijke gebruiker' naar de beleidsmakers van het ministerie van Buitenlandse Zaken die invloed hebben op de inhoud van beleid. Tot nu toe is steeds gesproken over beleidsmakers op het individuele niveau. In de praktijk wordt beleid echter zelden door één persoon bepaald en is het logischer om gebruik te beschouwen op organisatieniveau (Weiss, 1998). Concreet voor het onderzoek betekent dit dat de beleidskeuze(s) die gebaseerd is (of zijn) op de evaluatie via de formele lijnen in de organisatie moet(en) zijn genomen.

2.2 Qualitative Comparative Analysis als benadering en techniek

De vraag die in dit onderzoek centraal staat, is: welke combinaties van condities³ leiden tot

instrumenteel gebruik van een evaluatie; en welke tot de afwezigheid ervan? Qualitative Comparative Analysis (QCA) laat toe om een systematisch antwoord te vinden op deze hoofdvraag. Binnen het bestek van dit beknopte artikel kunnen wij niet in detail treden over QCA als ontologische benadering en familie van analysetechnieken. We verwijzen hiervoor naar basiswerken over de methode (Ragin, 1987, 2000, 2008; Rihoux & Ragin, 2009; Schneider & Wagemann, 2012). In het kader van dit artikel zouden we ons willen beperken tot het benadrukken dat QCA een methode is die relaties tussen condities beschouwt als sets en subsets en die zoekt naar condities die een noodzakelijke en/of voldoende voorwaarde zijn voor een uitkomst. Tijdens een QCA worden casussen systematisch vergeleken op combinaties van condities. Door Booleaanse minimalisatie worden deze combinaties van condities tot de kleinst mogelijke ‘verklarende paden’ voor de uitkomst teruggebracht (Marx & Dusa, 2011).

De methode is bijzonder geschikt om tegemoet te komen aan de in de inleiding genoemde tekortkomingen van bestaand onderzoek. Dit heeft vooral te maken met de welbepaalde benadering ten aanzien van causaliteit die QCA erop nahoudt. Deze benadering strookt goed met de causale relaties die we veronderstelden te vinden bij het complexe thema evaluatiegebruik. Causaliteit wordt opgevat als een samenspel van verschillende condities en niet als de afzonderlijke invloed van individuele condities op een uitkomst (conjuncturele causaliteit). Bovendien kan eenzelfde uitkomst veelal bereikt worden via verschillende combinaties van condities (equifinaliteit, Schneider & Wagemann, 2012: 5-6). Ten slotte wordt het uitblijven van een uitkomst niet per definitie bepaald door de omgekeerde oorzaak van het voorkomen van uitkomst (asymmetrische causaliteit, Schneider & Wagemann, 2012: 81). Met andere woorden vereist QCA een aparte analyse van de vraag welke condities het *aanwezig* zijn van evaluatiegebruik beïnvloeden en welke het *afwezig* zijn van gebruik.

Het eigenlijke onderzoek werd uitgevoerd in verschillende stappen. De eerste stap betrof het selecteren van de evaluaties die zouden worden betrokken in de analyse. Teneinde maximaal beroep te kunnen doen op het geheugen van medewerkers bij IOB, werden enkel deze evaluaties meegenomen die recent zijn uitgevoerd (maximaal drie jaar voorgaand aan het onderzoek). Alle geselecteerde evaluaties bevatten expliciet aanbevelingen die werden gedaan aan beleidsmedewerkers. Uiteindelijk werden achttien evaluaties meegenomen in de analyse. De evaluaties betreffen uiteenlopende thema’s op het gebied van handel, ontwikkelingssamenwerking en diplomatie.

Naast de selectie van evaluaties maakten we een selectie van condities waarvan we meenden dat deze instrumenteel gebruik van een evaluatie kunnen beïnvloeden. Hiertoe baseerden we ons in eerste instantie op de evaluatieliteratuur. Aanvankelijk resulteerde dit in een lange lijst van meer dan twintig condities. Gezien het vergelijkende karakter van een QCA zou een dergelijk lange lijst niet toelaten valide conclusies te trekken op basis van gelijkende combinaties van condities geïdentificeerd bij meerdere evaluaties. In plaats daarvan zou het resulteren in quasi-individuele verhalen per evaluatie en zodoende was een verdere selectie noodzakelijk. Enkel condities die voldoende varieerden in aan- en afwezigheid, werden meegenomen. Condities die vrijwel stabiel zijn over de evaluaties heen, hebben logischerwijze weinig verklarend potentieel. Hierbij vielen condities als ‘de kwaliteit van de evaluatie’ en ‘de deskundigheid van de onderzoeker’ af omdat deze in vrijwel alle gevallen als ‘hoog’ werden beschouwd door de geïnterviewde beleidsmedewerkers. Ook de in de literatuur belangrijke conditie ‘het betrekken van de gebruikers bij het onderzoek’ (Johnson et al., 2009: 382) viel af als conditie doordat er onvoldoende variatie bleek tussen de evaluaties. Op basis van het criterium van variabiliteit werden daarom zeven condities behouden. Van deze condities hebben we ons toegespitst op enkel die condities die op een betrouwbare wijze konden worden gemeten en ‘gekalibreerd’ (zie verder). Een conditie als ‘uitvoerbaar aanbevelingen’ werd bijvoorbeeld niet behouden in de analyse, omdat respondenten het veelal zelf moeilijk vonden om aan te geven of de aanbevelingen uitvoerbaar waren geweest. Ten slotte bleven er vier condities over met het grootste verklarend potentieel: (1) politiek gehalte van de evaluatie; (2) tijdigheid van de evaluatie; (3) nieuwe kennis in de evaluatie; en (4) interesse die werd getoond door de beleidsmakers. Met een verhouding van vier condities en achttien casussen beantwoordt het onderzoeksdesign aan ‘best practice’ geldend voor QCA-onderzoek (Marx & Dusa, 2011).⁴ Elk van deze condities lichten we verderop in het artikel toe.

Voor de eigenlijke verzameling van data over de condities en de uitkomst trianguleerden we verschillende methoden. Per evaluatie hielden we een semigestructureerd interview met minimaal één beleidsmaker en één betrokken evaluator. Deze inzichten werd aangevuld met informatie vergaard in een enquête uitgestuurd naar ongeveer drie andere beleidsmakers per casus. Daarnaast werd gebruikgemaakt van verschillende documenten met relevantie voor de evaluatie, waaronder bijvoorbeeld de beleidsbrief die naar de Tweede Kamer is gestuurd door de minister.

Met deze data voorhanden konden we overgaan tot de daadwerkelijke analyse. QCA maakt gebruik van Booleaanse algebra om casussen systematisch te kunnen vergelijken met elkaar. Dit veronderstelt dat elke casus, i.e. elke evaluatie, wordt 'vertaald' in een combinatie van gecodeerde uitkomst en condities als aanwezig (of hoog; sterk; ... code: 1) of afwezig (of laag; zwak; ... code: 0). Dergelijke vertaling in '1' en '0' wordt ook wel kalibrering genoemd.⁵ In deze studie deden we een beroep op de oorspronkelijke crisp set versie van QCA, waarbij we geen onderscheid hebben gemaakt tussen meer verfijnde 'degrees' van aanwezigheid of afwezigheid zoals fuzzy set QCA wel zou doen. Vanuit de beschikbare data kon niet precies worden ingeschat *in hoeverre* een conditie aanwezig was, wat een voorwaarde is om fuzzy-set-codering te gebruiken. Of een conditie eerder wel dan niet aanwezig was, konden we wel nauwkeurig bepalen. De QCA zelf werd uitgevoerd met behulp van het softwareprogramma R.⁶

3 Operationalisatie

3.1 Instrumenteel gebruik, de 'uitkomst' van deze studie

Zoals eerder beschreven betreft voorliggend onderzoek enkel het instrumenteel gebruik van evaluaties, eveneens conform de ambities van DAC en IOB. Instrumenteel gebruik bevat 'het gebruiken van een evaluatie bij het maken van beleidskeuzes' (eigen vertaling, Ledermann, 2012). Het gaat hierbij niet om 'window dressing', maar om een serieus gebruik van de bevindingen. Instrumenteel gebruik hoeft niet noodzakelijkerwijze opgeschreven en gedocumenteerd te zijn (Leviton & Hughes, 1981: 530). Dit maakt het onderzoek lastig: zwart-op-wit 'bewijs' voor gebruik is niet altijd te vinden, dus moet er worden vertrouwd op (mondelijke) informatie verkregen van de gebruikers zelf. Zoveel mogelijk zijn daarom de bevindingen getrianguleerd en is gebruik van een evaluatie door meerdere bronnen bevestigd, voordat deze effectief als 'gebruikt' werd beschouwd.

In dit onderzoek beschouwen we een evaluatie als instrumenteel gebruikt (code: 1) wanneer er ten minste één grote beleidskeuze significant beïnvloed is door de evaluatie. Grote beleidskeuzes betreffen bijvoorbeeld: het besluit om beleid (deels) te stoppen of door te zetten; een strategische wijziging in het beleid met consequenties op het operationele niveau; of belangrijke wijzigingen in de financiering. De afwezigheid van evaluatiegebruik (code: 0) betreft dan logischerwijze de eis dat er geen enkele belangrijke beleidskeuze significant door de evaluatie is beïnvloed. We erkennen de mogelijkheid dat er in dat geval wel meer mineure zaken werden aangepast op basis van de evaluatie, bijvoorbeeld de beslissing om (al gemaakte) beleidskeuzes expliciet op papier te zetten.

3.2 Verklarende condities

Het politieke gehalte van het evaluatie-onderwerp

De evaluatie heeft een hoog politiek gehalte (code: 1) wanneer het onderwerp van de evaluatie politiek gevoelig is of hoog op de politieke agenda staat.

De literatuur is niet eensgezind over de invloed van het politieke gehalte van evaluatieonderwerpen op gebruik, wat de relevantie van deze conditie nog meer het onderzoeken waard maakt. Ledermann (2012) laat zien dat een politiek geladen onderwerp een positief effect heeft op evaluatiegebruik. Haar verklaring hiervoor is dat politiek geladen onderwerpen beter onderbouwd dienen te worden om door het parlement geloodst te worden. Barrios (1986: 111) daarentegen beargumenteert dat een evaluatie over een minder politiek onderwerp juist eerder gebruikt wordt. Als er geen politieke lijn is waarop beleidsmakers terug kunnen (of moeten) vallen, dan zullen ze zoeken naar andere bronnen,

zoals evaluaties, om hun beleid op te baseren. Aan alle respondenten is gevraagd of het onderwerp van de evaluatie als politiek gevoelig kan worden beschouwd dan wel hoog op de politieke agenda staat. Wanneer uit de verschillende antwoorden het eenduidige beeld naar voren kwam dat het evaluatieonderwerp (in ieder geval grotendeels) politiek gevoelig of hoog op de agenda stond, werd besloten dat de evaluatie een hoog politiek gehalte had.

Tijdigheid

De evaluatie is tijdig (code: 1) wanneer de beleidsafdeling werkt aan een nieuw beleid of aan grote beleidsveranderingen tegelijk met de fase van datacollectie van de evaluatie en/of het schrijven van het evaluatierapport. Een nieuw beleid of grote wijzigingen behelst een complete verandering van beleid, evenals belangrijke veranderingen in focus, omvang en doelen van beleid.

Verschillende onderzoeken hebben laten zien dat de tijdigheid van een evaluatie van belang is voor het gebruik (Bober & Bartlett, 2004: 377; Boyer & Langbein, 1991: 527; Rockwell et al., 1990; Shea, 1991: 107). In deze onderzoeken wordt 'tijdigheid' doorgaans gemeten door aan de potentiële gebruikers te vragen: 'Kwam de evaluatie op een ideaal tijdstip, en niet te vroeg of te laat?' Weinig verrassend worden evaluaties die op een ideaal tijdstip komen, meer gebruikt dan evaluaties die te vroeg of te laat zijn. Dit vertelt ons echter niet wanneer een evaluatie als op tijd kan worden beschouwd.

In dit onderzoek hebben we ervoor gekozen tijdigheid concreet te definiëren. We beschouwen een evaluatie als tijdig wanneer het proces van de evaluatie gelijk loopt aan het proces van het vormen van nieuw beleid. Het idee hierachter is dat er meer mogelijkheden zijn voor beïnvloeding en meer momenten om resultaten over te dragen wanneer een evaluatieproces gelijk loopt aan het proces van beleidsformulering, dan als het evaluatieproces daarvoor of daarna loopt.

Nieuwe kennis

De evaluatie bevat nieuwe kennis (code: 1) wanneer de voornaamste beleidsmakers aangeven dat de evaluatie kennis bevat over het beleid of het beleidsonderwerp die nieuw voor hen was.⁷

De literatuur geeft geen eenduidig antwoord op de vraag of een evaluatie nieuwe kennis voor de beleidsmakers moet bevatten opdat deze gebruikt zal worden. Ledermann (2012: 173) laat zien dat nieuwe kennis een noodzakelijke conditie is voor instrumenteel gebruik wanneer deze een politiek conflictueus thema betreft. Als de kennis niet nieuw is, is het waarschijnlijker, volgens Ledermann, dat de evaluatie symbolisch, en dus niet instrumenteel, gebruikt wordt. Johnson et al. (2009: 385), in hun literatuurreview, concluderen dat kennis in een evaluatie bevestiging nodig heeft van externe informatie. Hieruit valt af te leiden dat evaluaties die nieuwe kennis aanleveren, juist eerder aan de kant kunnen worden gelegd in afwachting van extra evidentie ter zake. Opnieuw maakt de onduidelijkheid van het exacte verband dit een interessante conditie om te onderzoeken. Of een evaluatie op het moment van oplevering nieuwe kennis aanreikte, werd bevraagd in de interviews en de survey met beleidsmakers.

Getoonde interesse van de beleidsmakers

De evaluatie wordt als interessant voor de beleidsmakers (code: 1) beschouwd als de voornaamste beleidsmakers (a) de bevindingen gecommuniceerd hebben aan de relevante beleidsmakers⁸ tijdens een vergadering of via een presentatie van de evaluatoren; of (b) ten minste één vraag naar voren brachten aan het begin van de evaluatie waarop ze hoopten een antwoord krijgen in de evaluatie.

Een aantal studies hebben aangetoond dat interesse van de beleidsmakers een positief effect heeft op het gebruik van evaluaties (Johnson et al., 2009; Leviton & Hughes, 1981; Patton et al., 1977; Preskill et al., 2003). Daarom hebben we ook deze conditie meegenomen in het huidige onderzoek.

De kans op sociaal wenselijke antwoorden bij het meten van interesse is groot wanneer men expliciet vraagt aan beleidsmakers of ze geïnteresseerd zijn. Daarom hanteerden we een proxy om de interesse betrouwbaar in beeld te brengen. Bij dit onderzoek wordt aangenomen dat er sprake is van interesse als de beleidsmakers de evaluatoren aan het begin van het evaluatieproces specifieke (onderzoeks)vragen meegeven. Daarnaast kunnen de beleidsmakers interesse tonen door de resultaten van het onderzoek breed te bespreken binnen hun afdeling. Opnieuw hebben we via interviews en de survey onderzocht of er effectief aan één of meer van de voorwaarden voor interesse van beleidsmakers werd voldaan.

3.3 Binaire vertaling van de bestudeerde evaluaties

Wanneer we bovenstaande operationalisatieprincipes toepassen, kunnen we de door ons geobserveerde evaluaties vertalen in een combinatie van binair vertaalde condities en de uitkomst 'evaluatiegebruik'. De onderstaande 'waarheidstabel' geeft alle logische combinaties van condities weer en de uitkomst die empirisch is waargenomen bij die combinatie. Waar we een bepaalde combinatie niet empirisch konden vaststellen, is een vraagteken geplaatst in de uitkomstkolom. In dit onderzoek zijn geen aannames gedaan met betrekking tot deze zogenaamde 'logical remainders' en deze combinaties zijn zodoende niet meegenomen in de analyse.

In totaal observeerden we achttien evaluaties, waarvan er vijf als 'gebruikt' kunnen worden beschouwd.⁸ De ongelijke verdeling van data, zeker met betrekking tot de uitkomstconditie, maakt de analyse vrij lastig. De precieze gevolgen van deze vorm van 'skewedness' zijn in de literatuur nog onderbelicht (Schneider & Wagemann, 2012: 248). In elk geval kan worden gesteld dat de bevindingen niet zonder meer als harde conclusies mogen worden beschouwd of zomaar mogen worden veralgemeniseerd. Niettemin levert de analyse wel een aantal interessante indicaties op. We bespreken deze in de volgende sectie.

Tabel 2 Waarheidstabel

Alle mogelijke combinaties en eventueel waargenomen uitkomst						
Rijnr.	Politiek gehalte	Tijdigheid	Kennis	Interesse	Uitkomst	Aantal casussen
1	0	0	0	0	0	3
2	0	0	0	1	0	1
3	0	0	1	1	0	1
4	1	0	0	0	0	2
5	1	0	0	1	0	1
6	1	0	1	1	0	1
7	1	1	0	0	0	1
8	1	1	0	1	0	4
9	0	0	1	0	?	0
10	0	1	0	0	?	0
11	0	1	0	1	?	0
12	0	1	1	0	?	0
13	1	0	1	0	?	0
14	1	1	1	0	?	0
15	0	1	1	1	1	1
16	1	1	1	1	1	3

De tabel bevat één rij met een ‘logical contradiction’: dezelfde combinatie condities levert drie evaluaties op waar geen gebruik plaatsvond en één evaluatie waarin wel sprake was van gebruik (rij 8). De consistentiescore voor deze rij kwam daarmee op 0.75. De consistentiescore nodig om meegenomen te worden in de analyse is op 0.7 gezet. Met 0.75 is rij 8 meegenomen als ‘niet gebruikt’ (code 0).

4 Resultaten

4.1 Analyse van condities die noodzakelijk zijn voor evaluatiegebruik

We kunnen een conditie als strikt noodzakelijk beschouwen wanneer de conditie aanwezig (of afwezig) is in alle evaluaties die instrumenteel gebruikt werden (of niet). In tabel 3 staan de resultaten voor de analyse van noodzakelijkheid. De cijfers in de cellen geven weer voor hoeveel evaluaties een verband werd geobserveerd voor elk van de vier condities en de uitkomst.

Tabel 3 Noodzakelijkheid

		Uitkomst	
		Niet aanwezig	Aanwezig
Politiek gehalte	Niet aanwezig	5	1
	Aanwezig	8	4
		Uitkomst	
		Niet aanwezig	Aanwezig
Kennis	Niet aanwezig	11	1
	Aanwezig	2	4
		Uitkomst	
		Niet aanwezig	Aanwezig
Tijdigheid	Niet aanwezig	9	0
	Aanwezig	4	5
		Uitkomst	
		Niet aanwezig	Aanwezig
Interesse	Niet aanwezig	6	0
	Aanwezig	7	5

Voor een juiste interpretatie van de bevindingen worden in tabel 4 eveneens de consistentie- en dekkingsscores van de noodzakelijkheidsanalyse weergegeven. Een hogere consistentiescore betekent dat een uitkomst vaker in combinatie voorkomt met de conditie. Bij een consistentiescore van 1 is de conditie altijd aanwezig, wanneer de uitkomst aanwezig is. Uit tabel 4 blijkt dat de condities tijdigheid en interesse volledig consistent zijn. De overige twee condities, politiek en kennis, scoren een ‘0.8’, wat betekent dat deze in tachtig procent van de gebruikte evaluaties aanwezig bleken. Schneider en Wagemann (2012: 143) adviseren om enkel condities die hoger dan ‘0.9’ scoren als noodzakelijk te beschouwen. Dat betekent dat politiek en kennis niet als noodzakelijk gekwalificeerd worden. Getoonde interesse vanuit de beleidsmakers en tijdigheid van de evaluatie blijken noodzakelijke factoren te zijn voor het gebruik van evaluaties. Ons onderzoek bevestigt het belang dat aan deze factoren werd gehecht in eerdere studies. Daarenboven wordt

aangetoond wat een tijdig moment is om de evaluatie uit te voeren: die kan als optimaal worden beschouwd wanneer het evaluatieonderzoek tegelijk loopt met het formuleren van nieuw beleid door de beleidsmakers. Dit laat zich goed verklaren: er dienen zich veel mogelijkheden aan tot kennisoverdracht en het is relatief makkelijk om nieuwe inzichten in beleid te vertalen. Bovendien hebben beleidsmakers juist op dat moment behoefte aan de inzichten die de evaluatie oplevert.

Naast de consistentiescore moet ook worden gekeken naar de zogenaamde dekkingsscore. Deze term is wat misleidend; de term ‘trivialiteit’ zou beter zijn (vgl. Schneider & Wagemann, 2012: 144). Als deze score namelijk laag is, is het aantal casussen waarbij de uitkomst aanwezig is én de conditie aanwezig is, klein in vergelijking met alle casussen waarbij de conditie aanwezig is. Bij een score van ‘1’ is de conditie altijd in combinatie met de uitkomst aanwezig, bij ‘0’ komt de conditie helemaal niet voor in combinatie met de uitkomst. Bij een lage score differentiëren de condities niet in zeer grote mate tussen gebruik en niet-gebruik van evaluaties. De dekkingsscores in deze analyse zijn behoorlijk laag. Omdat tijdigheid en interesse beide een consistentiescore van 1 hebben, zijn deze twee met name interessant. Tijdigheid scoort ‘0.56’. Dit betekent dat van de negen evaluaties in de set ‘tijdige evaluaties’, er ook vijf daadwerkelijk instrumenteel gebruikt werden. Interesse scoort ‘0.42’: van de twaalf evaluaties in de set ‘evaluaties waarin interesse is getoond’ zijn er ook vijf in de set ‘gebruikte evaluaties’. Hieruit valt op te maken dat tijdigheid en interesse weliswaar noodzakelijk zijn, maar dat hun aanwezigheid zeker niet direct leidt tot gebruik.

Tabel 4 Consistentie- en dekkingsscores voor noodzakelijkheid

	Consistentiescore uitkomst	Dekkingsscore uitkomst	Consistentiescore niet-uitkomst	Dekkingsscore niet-uitkomst
Politiek gehalte	0.8	0.33	0.62	0.67
Tijdigheid	1	0.56	0.39	0.44
Kennis	0.8	0.67	0.15	0.33
Interesse	1	0.42	0.58	0.58

In tabel 3 en 4 zijn ook de scores terug te vinden voor de analyse van noodzakelijkheid van de afwezigheid van de uitkomst. Uit tabel 4 blijkt echter dat geen van de condities noodzakelijk is voor de afwezigheid van instrumenteel gebruik. Voor praktische doeleinden kan het een betreurenswaardige conclusie zijn dat er geen noodzakelijk factoren zijn gevonden. Er is geen welbepaalde sleutelfactor die verklaart waarom evaluaties niet werden benut.

4.2 Analyse van condities die voldoende zijn

Een conditie kan als voldoende voorwaarde voor een uitkomst worden gekwalificeerd wanneer er geen enkele evaluatie is waar de conditie wel voorkomt, maar de uitkomst niet. Oftewel: er mag geen casus zijn waarbij de conditie een ‘1’ scoorde, maar de uitkomst een ‘0’.

In tabel 3 is te zien dat geen van de condities op zichzelf voldoende is voor de uitkomst. De analyse richt zich daarom verder volledig op *combinaties* van condities die voldoende zijn voor de aan- of afwezigheid van de uitkomst, zogenaamde paden die leiden tot de uitkomst. Vanuit de waarheidstabel (tabel 2) is door middel van ‘logische minimalisering’ een oplossing (‘solution term’) verkregen voor de aanwezigheid van de uitkomstconditie. Logische minimalisering volgt de ‘één-verschil-regel’ (Baumgartner, 2012) en is gebaseerd op de assumptie dat een conditie geen echte verklarende waarde heeft en bijgevolg overbodig is wanneer twee combinaties voor slechts één conditie een verschillende waarde krijgen terwijl ze dezelfde uitkomst vertonen. Wanneer je deze regel iteratief herhaalt tot er geen verdere vereenvoudiging meer mogelijk is, kom je tot de meest zuinige (‘parsimonious’) paden die geassocieerd kunnen worden met een welbepaalde uitkomst. Het

resultaat van deze logische minimalisering voor onze studie is weergegeven in tabel 5. In formele termen is de oplossing leidend naar evaluatiegebruik:

TIJDIGHEID * KENNIS * INTERESSE → UITKOMST⁹

Tabel 5 Resultaten van de analyse van voldoende condities voor de aanwezigheid van de uitkomst

	Politiek	Tijdigheid	Kennis	Interesse	Consistentie	Ruwe dekkingsgraad
1:		•	•	•	1	0.8

Consistentie van de gehele oplossing: 1 / Dekkingsgraad van de gehele oplossing: 0.8
Configuratie leidend tot de aanwezigheid van de uitkomst (UIT): instrumenteel gebruik van de evaluatie. • staat voor de aanwezigheid van een verklarende conditie in de oplossing; ø staat voor de afwezigheid van de conditie. NB: de unieke dekkingsgraad is niet getoond, aangezien die gelijk is aan de ruwe dekkingsgraad.

De logische minimalisering leert ons dat een combinatie van drie factoren voldoende is om de uitkomst tot stand te brengen: tijdigheid én nieuwe kennis én interesse. Deze oplossing heeft een consistentiescore van '1'. De dekkingscore is '0.8', aangezien vier van de vijf casussen waar de uitkomst aanwezig is, door dit pad worden verklaard.

De verklaring van het gevonden pad spreekt voor zich: tijdigheid geeft een goede mogelijkheid tot gebruik, nieuwe kennis laat zien wat er in het beleid kan veranderen en interesse laat zien dat men gemotiveerd was om de evaluatie te gebruiken. Of een evaluatie al dan niet een politiek geladen thema betreft, lijkt dan niet veel uit te maken. Dit kan mogelijk worden verklaard doordat meestal enkel een deel van de evaluatie echt politiek gevoelig van aard is, en minder politiek getinte aanbevelingen wel geïmplementeerd kunnen worden.

In tabel 6 staat vervolgens de oplossing voor de *afwezigheid* van evaluatiegebruik. Deze oplossing geeft drie verschillende paden, die allemaal een consistentiescore van '1' hebben. In formele termen is de oplossing:

tijdigheid * kennis + tijdigheid * INTERESSE + POLITIEK GEHALTE * kennis * interesse
→ uitkomst

Tabel 6 Resultaten van de analyse van voldoende condities voor de afwezigheid van de uitkomst

	Politiek	Tijdigheid	Kennis	Interesse	Consistentie	Ruwe dekkingsgraad	Unieke dekkingsgraad
1:		ø	ø		1	0.54	0.23
2:		ø		•	1	0.31	0.15
3:	•		ø	ø	1	0.23	0.08

Consistentie van de gehele oplossing: 1 / Dekkingsgraad van de gehele oplossing: 0.77
Configuratie leidend tot de afwezigheid van de uitkomst (uit): instrumenteel gebruik van de evaluatie. • staat voor de aanwezigheid van een verklarende conditie in de oplossing; ø staat voor de afwezigheid van de conditie.

De tabel rangschikt de verschillende paden naar empirische relevantie via de dekkingscores (0.77 voor de paden samen). Het eerste pad betreft de aanwezigheid van nieuw aangeboden kennis en een suboptimale tijdigheid. Het spreekt voor zich dat als een evaluatie niet tijdig is, het niet evident is voor de beleidsmakers om veranderingen door te voeren: de gelegenheid daarvoor zou speciaal gecreëerd moeten worden. Als er dan evenmin nieuwe kennis wordt aangeleverd via de evaluatie, is er dan ook weinig prikkel om die gelegenheid

te creëren.

Het derde pad is ook niet zo ingewikkeld om te verklaren. Zonder kennis en interesse is de stimulus om de evaluatie te gebruiken sowieso beperkt. Indien de evaluatie dan eveneens een politiek geladen karakter heeft, is de kans nog groter dat beleidsmakers er hun nek niet voor willen uitsteken. Onze bevindingen lijken dus te stroken met eerder onderzoek waarin werd aangegeven dat politiek getinte evaluaties minder gebruikt worden (Barrios, 1986: 111).

De laatste mogelijkheid, het tweede weergegeven pad, is echter wel lastig te verklaren. Dit betreft de evaluaties die gekenmerkt worden door een suboptimale timing, maar waar in feite bij de beleidsmakers wel interesse was voor de bevindingen. Dat een gebrek aan tijdigheid kan leiden tot geen gebruik is niet vreemd, maar dat de aanwezigheid van interesse toch niet leidt tot het gebruik van bevindingen valt moeilijker toe te lichten. Ook in de literatuur en in de interviews die we hielden voor dit onderzoek, konden we hiertoe geen mogelijke causale mechanismen vinden. Een meer gedetailleerde screening van de betrokken evaluaties leert ons dat mogelijk een andere conditie verklarend is. Voor twee van de vier evaluaties die gedekt worden door dit pad, bleek dat het hogere management (en dus niet de direct verantwoordelijke beleidsmaker) weinig interesse toonde in de evaluatie. Het belang van de houding van het management werd eveneens al aangetoond in eerdere studies over evaluatiepraktijken (Pattyn, 2014). Helaas beschikken we niet over gegevens over de houding van het hoger management van alle achttien evaluaties om ook deze conditie te testen. We merken verder op dat de overige twee evaluaties waarvoor dit pad geldt, ook kunnen worden uitgelegd door het eerder beschreven pad 'geen tijdigheid en het ontbreken van nieuwe kennis'. Dit lijkt ons een betere verklaring voor het gebrek aan gebruik, dan de aanwezigheid van interesse. De uitdagingen in het verklaren van de associatie tussen het ontbreken van tijdigheid en de aanwezigheid van interesse, en de mogelijke rol van een extra verklarende factor in dezen, wijzen ons op het belang van een gegronde theoretische basis alvorens over te gaan tot een QCA-analyse. Anderzijds helpt de QCA ons ook om het theoretische model verder te verfijnen.

5 Slotbeschouwing

Met deze studie hadden we de ambitie om zowel voor de theorie als voor de praktijk nuttige inzichten te verschaffen in de factoren die invloed hebben op wel of geen gebruik van evaluaties. Op theoretisch vlak hebben we getracht vanuit een systematische invalshoek bouwstenen aan te leveren voor een verdere onderzoeksagenda rond het thema 'evaluatiegebruik'. De conjuncturele assumptie onderliggend aan QCA hielp ons verder te kijken dan louter de invloed van individuele factoren op evaluatiegebruik en complementeerde dit met een focus op de combinaties van factoren die kunnen leiden tot gebruik. De analyse bevestigde dat het verklaren van evaluatiegebruik een complex verhaal is, waar meerdere causale verhalen aan ten grondslag liggen. Verder onderzoek spitst zich idealiter nog toe op andere condities waarvan we menen dat deze het theoretisch model nog kunnen versterken. Uit onze analyse bleek de rol van het hoger management bijvoorbeeld nog het nader onderzoeken waard.

Voor de praktijk van beleidsmakers en evaluatoren benadrukken we het belang van tijdigheid en interesse voor een evaluatie. Onze analyse laat zien dat professionals die op deze factoren inzetten, een gunstig klimaat creëren voor het gebruik van evaluaties. Wat betreft tijdigheid lijkt het van belang de evaluatie te laten sporen met het schrijfwerk van een beleidsafdeling aan nieuw beleid of grote beleidsveranderingen. Goede anticipatie en een sterke institutionalisering van het evaluatieproces zijn hiertoe cruciaal. Of het thema van de evaluatie een sterk politieke gevoeligheid kent, is minder belangrijk. Mits er sprake is van de juiste omgevingscondities, kunnen ook dergelijke evaluaties sterk instrumenteel worden benut.

Literatuur

Alkin, M.C. (Ed.). (2004). *Evaluation roots*. Thousand Oaks: Sage.

Alkin, M.C., & Taut, S.M. (2003). Unbundling evaluation use. *Studies in Educational Evaluation*, 29, 1-12.

- Azzam, T. & Levine, B. (2015). Politics in evaluation: Politically responsive evaluation in high stakes environments. *Evaluation and Program Planning*, 53, 44-56.
- Bouterse, M. (2016). *Explaining evaluation use: A Qualitative Comparative Analysis of factors influencing instrumental use of evaluations*. Masterscriptie. Universiteit Leiden.
- Balthasar, A. (2006). The effects of institutional design on the utilization of evaluation: Evidenced using Qualitative Comparative Analysis (QCA). *Evaluation*, 12(3), 353-371.
- Barrios, N.B. (1986). *Utilization of evaluation information: A case study approach investigating factors related to evaluation utilization in a large state agency*. Proefschrift. Florida State University: ProQuest Dissertations Publishing.
- Basurto, X., & Speer, J. (2012). Structuring the calibration of qualitative data as sets for qualitative comparative analysis (QCA). *Field Methods*, 24(2), 155-174.
- Baumgartner, M. (2012). Detecting causal chains in small-n data. *Field Methods*, 25, 3-24.
- Bober, C.F., & Bartlett, K.R. (2004). The utilization of training program evaluation in corporate universities. *Human Resource Development Quarterly*, 15(4), 363-383.
- Boyer, J.F., & Langbein, L.I. (1991). Factors influencing the use of health evaluation research in congress. *Evaluation Review*, 15(5), 507-532.
- Development Assistance Committee. (1991). *Principles for the evaluation of development assistance*. Paris.
- Dye, T.R. (1972). *Understanding public policy*. New Jersey: Prentice Hall.
- Henry, G.T., & Mark, M. (2003). Toward an agenda for research on evaluation. *New Directions for Evaluation*, 97, 69-80.
- Johnson, K., Greenesid, L.O., Toal, S.A., King, J.A., Lawrenz, F., & Volkov, B. (2009). Research on evaluation use: A review of the empirical literature from 1986 to 2005. *American Journal of Evaluation*, 30(3), 377-410.
- Kirkhart, K.E. (2000). Reconceptualization evaluation use: An integrated theory of influence. *New Directions for Evaluation*, 88, 5-22.
- Korsten, A.F.A. (2013). Wat evaluatoren van kritiek kunnen leren. *Beleidsonderzoek Online*, december. DOI: 10.5553/Beleidsonderzoek.000030.
- Korsten, A.F.A., & Van der Meer, A.D. (2014). Beleidsonderzoek benutten. *Beleidsonderzoek Online*, februari. DOI: 10.5553/Beleidsonderzoek.000038.
- Ledermann, S. (2012). Exploring the necessary conditions for evaluation use in program change. *American Journal of Evaluation*, 33(2), 159-178.
- Leviton, L.C., & Hughes, E.F.X. (1981). Research on the utilization of evaluations: A review and synthesis. *Evaluation Review*, 5(4), 525-548.
- Marx, A., & Dusa, A. (2011). Crisp-set qualitative comparative analysis (csQCA), contradictions and consistency benchmarks for model specification. *Methodological Innovations Online*, 6(2), 103-148.
- Ministerie van Buitenlandse Zaken. (2009) *Evaluatiebeleid en richtlijnen voor evaluaties*. Den Haag. Op 6 april 2016 ontleend aan <https://www.rijksoverheid.nl/documenten/brochures/2009/10/01/evaluatiebeleid-en->

richtlijnen-voor-evaluaties.

Patton, M.Q., Grimes, P.S., Guthrie, K.M., Brennan, N.J., French, B.D., & Blyth, D.A. (1977). In search of impact: An analysis of the utilization of federal health evaluation research. In C. Weiss (Ed.), *Using social research in public policy making* (pp. 141-163). Lexington: Lexington Books.

Pattyn V. (2014). Why organisations (do not) evaluate? Explaining evaluation activity through the lens of configurational comparative methods. *Evaluation: The International Journal of Theory, Research and Practice*, 20(3), 348-367.

Pattyn, V., Molenveld, A., & Befani, B. (2017). Qualitative Comparative Analysis as an evaluation tool: Lessons from an application in development cooperation. *American Journal of Evaluation*, Online First.

Preskill, H., Zuckerman, B., & Matthews, B. (2003). An exploration study of process use: Findings and implications for future research. *American Journal of Evaluation*, 24(4), 423-442.

Quaranta, M., & Schneider, C.Q. (2013). *How to... Use software for set-theoretic analysis: Online appendix to Carsten Q. Schneider and Claudius Wagemann (2012) 'Set-theoretic methods for the social sciences'*. Cambridge University Press. Op 31 augustus 2016 ontleend aan www.cambridge.org/US/download_file/134448/.

Ragin, C.C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies*. Berkeley: University of California Press.

Ragin, C.C. (2000). *Fuzzy set social science*. Chicago, IL: University Chicago Press.

Ragin, C.C. (2008). *Redesigning social inquiry: Fuzzy sets and beyond*. Chicago, IL: University Chicago Press.

Rihoux, B., & Ragin, C. (2009). *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*. Thousand Oaks, CA: Sage.

Rockwell, S.K., Dickey, E.C., & Jasa, P.J. (1990). The personal factor in evaluation use: A case study of steering committee's use of a conservation tillage survey. *Evaluation and Program Planning*, 13, 389-394.

Schneider, C.Q., & Wagemann, C. (2012). *Set-theoretic methods for the social sciences: A guide to qualitative comparative analysis*. Cambridge, UK: Cambridge University Press.

Shea, M.P. (1991). *Program evaluation utilization in Canada and its relationship to evaluation process, evaluator and decision context variables*. Proefschrift. University of Windsor, Canada: ProQuest Dissertations Publishing.

Van Hoesel, P. (2017). Beleidsevaluatie als ritueel. *Beleidsonderzoek Online*, januari. DOI: 10.5553/BO/221335502017000001001.

Weiss, C.H. (1979). The many meanings of research utilization. *Public Administration Review*, 39(5), 426-431.

Weiss, C.H. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation*, 19(1), 21-33.

Noten

* We willen het ministerie van Buitenlandse Zaken en de directie Internationaal Onderzoek en Beleidsevaluatie (voorheen Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie)

in het bijzonder bedanken voor het bieden van de gelegenheid voor dit onderzoek. Ook willen we de vele respondenten bedanken die bij hebben gedragen door hun kennis en inzichten te delen.

1 Het onderzoek dat wordt gepresenteerd in dit artikel, betreft een scriptieonderzoek dat door Marjolein Bouterse werd uitgevoerd als onderdeel van haar research master in politieke wetenschappen aan Universiteit Leiden (Bouterse, 2016). Valérie Pattyn trad op als scriptiebegeleider.

2 Sinds juli 2017 de directie Internationaal Onderzoek en Beleidsevaluatie.

3 Qualitative Comparative Analysis hanteert een eigen jargon. Zo worden onafhankelijke variabelen ‘conditions’ of condities genoemd, en wordt naar de afhankelijke variabele verwezen als ‘outcome’ of ‘uitkomst’.

4 Bij meer condities ten opzichte van dit aantal casussen zou de verklarende waarde van de bevindingen snel kleiner worden. Het risico bestaat dan dat er onvoldoende minimalisatie mogelijk is en elke evaluatie een eigen verklarend pad voor de uitkomst heeft. Dit leert ons vervolgens niets in algemene zin over evaluatiegebruik.

5 In deze studie werd de kalibratiemethode van Basurto en Speer (2012) in aangepaste vorm gebruikt.

6 De volgende ‘R-packages’ voor QCA zijn gebruikt om de ‘truth table’ en de configuraties te verkrijgen: ‘QCA’, ‘QCA3’ en ‘SetMethods’. De online appendix van Quaranta en Schneider (2013) bij het boek van Schneider en Wagemann (2012) is bijzonder nuttig geweest in dit proces.

7 Het kan hier natuurlijk ook om één beleidsmaker gaan, als er geen andere beleidsmakers zijn die zo specifiek op hetzelfde onderwerp werken. De voornaamste beleidsmaker wordt gedefinieerd als: de beleidsmaker die de directe verantwoordelijkheid heeft voor het formuleren (en mogelijk ook implementeren) van het beleid. Dit is niet gelijk aan alle relevante beleidsmakers, waarmee alle beleidsmakers die direct of indirect betrokken zijn bij het beleidsonderwerp, worden bedoeld.

8 Eén gebruikte evaluatie is in de waarheidstabel lastig terug te vinden, omdat het gaat om één van de evaluaties in rij 8. Omdat drie van de vier niet gebruikte evaluaties betreft, staat de uitkomst van de rij als geheel op ‘niet gebruikt’. Zie ook de toelichting onder tabel 2.

9 In formele termen geldt dat aanwezigheid wordt gemarkeerd door het gebruik van hoofdletters en afwezigheid door het gebruik van kleine letters. Het *-teken kan als ‘en’ worden gelezen, terwijl het +-teken als ‘of’ kan worden gelezen.