

Beleidsevaluaties tussen methode en praktijk: Naar een meer realistische evaluatiebenadering

Valérie Pattyn & Stefan Verweij

Gepubliceerd in: *Burger, Bestuur & Beleid*

1. Introductie

Het idee dat evaluatie kan bijdragen aan het verbeteren van beleid is niets nieuws onder de zon. Al in de jaren '80 van de vorige eeuw bloeide evaluatie op met de bedoeling beleidsformulering en beleidsuitvoering te informeren, en in het hedendaagse 'evidence-based' tijdperk is evaluatie meer dan ooit een bloeiende praktijk.ⁱ Nederland volgde historisch weliswaar een ander en sneller traject dan België, waar evaluatie slechts na de millenniumwisseling algemeen ingang vond.ⁱⁱ De potentie van evaluatie is tegenwoordig niettemin grotendeels (h)erkend en wordt gevat in adagia als 'leren van evalueren'. Er worden met regelmaat evaluatiestudies uitgevoerd, maar opvallend is dat deze in toenemende mate ook het gebruik van evaluatieonderzoek zelf betreffen. Recente Nederlandse voorbeelden betreffen rapporten in opdracht van het Ministerie van Buitenlandse Zaken en door het Planbureau voor de Leefomgeving en het Kennisinstituut voor Mobiliteitsbeleid.ⁱⁱⁱ In Vlaanderen werd gebruik van evaluaties onder meer opgenomen in de MIRA-BE beleidsevaluatierapporten in het domein van milieu^{iv}, of in onderzoek omtrent de praktijk van toekomstverkenningen.^v Uit dergelijke publicaties blijkt dat de leerpotentie van evaluaties zeker wordt (h)erkend, maar dat uitkomsten van evaluatieonderzoek niet altijd worden benut of slechts selectief, onder andere doordat uitkomsten niet aansluiten bij de geëvalueerde praktijk.^{vi} Dat laatste wijst op een fundamentele wanverhouding tussen methode en praktijk die door Kerseboom als volgt wordt verwoordt: *de nadruk op 'what works' verwordt in de praktijk vaak tot het onderzoeken van 'what can be measured'*.^{vii} Er is vaak een discrepantie tussen ideeën en verwachtingen die beleidsmakers, politici en vaak genoeg ook evaluatoren, hebben ten aanzien van de evaluatiemethodiek en de complexe beleidspraktijk die wordt geëvalueerd.^{viii} In dit verband is een veelgehoorde frustratie bij politici en andere opdrachtgevers van beleidsevaluaties dat evaluatoren veelal *niet in staat zijn om ondubbelzinnige antwoorden te geven op de vraag waarom programma's falen of succesvol zijn*.^{ix} Leren van evalueren wordt door deze mismatch frequent gehypothekeerd. In deze bijdrage willen wij de vinger leggen op deze discrepantie en een lans breken voor een meer 'realistische' benadering voor beleidsevaluatie. Het doel is om methode en praktijk weer een beetje dichter bij elkaar te brengen en zo bij te dragen aan het debat over hoe beleidsmakers het best kunnen leren van evaluaties. We concentreren ons op de zogenaamde administratieve evaluaties, uitgevoerd in opdracht van of door een overheid. We laten de vormen van juridische evaluatie (door rechterlijke instanties) of politieke evaluaties (door burgers, via verkiezingen bijvoorbeeld) buiten beschouwing.^x

Dit artikel is verder als volgt opgebouwd. In de eerstvolgende paragraaf zetten we uiteen waarin de wanverhouding tussen methode en praktijk gegrond is. Daarna bespreken we met behulp van enkele illustraties *realistic evaluation* (RE)^{xi} als evaluatiebenadering om methode en praktijk te convergeren en zien we dat dit ook in de beleidspraktijk langzaam aan voeten in de aarden krijgt. Ter wille van het betoog maken we daarbij een enigszins gestereotypeerde vergelijking tussen 'de' klassieke en 'de' realistische evaluatiebenadering. Vervolgens proponeren we 'Qualitative Comparative Analysis' (QCA) als veelbelovende methodiek om beleidslessen te trekken uit realistische evaluaties.

2. Discrepantie

Ten grondslag aan het implementeren van beleid ligt het abstracte idee dat er voor en na de implementatie een verandering is op die condities die het beleid beoogt te beïnvloeden. Evaluatie is vervolgens gericht op het bepalen van de mate waarin het beleid de veroorzaker is geweest van de

veranderde condities, als die zijn opgetreden. De zogenaamde attributievraag staat in deze centraal. Deze klassieke ‘gouden standaard’ van evaluatie vereist dat er (1) een nulmeting is uitgevoerd van de beginsituatie (ex-ante meting) die kan worden vergeleken met (2) de eindsituatie (ex-post meting). En om er zeker van te zijn dat de verandering in de beoogde condities daadwerkelijk toe te schrijven is aan het geïmplementeerde beleid is het nodig dat (3) de beleidssituatie (de ‘policy on’ situatie) wordt vergeleken met een zo gelijkaardig mogelijke setting waarin het beleidsinitiatief niet plaatsvond (de ‘policy off’ situatie). Dit experimenteel design is afgebeeld in figuur 1.^{xii}

(figuur 1: experimenteel design)

De aantrekkingskracht van dit idee ligt in de ogenschijnlijke eenvoud ervan: als alle condities minus het beleid constant blijven dan kan precies worden vastgesteld wat het effect van het beleid is geweest. Hoewel enigszins gechargeerd is dit nog altijd vaak de benadering die is terug te vinden in beleidsevaluaties. Uitspraken met de vorm ‘het effect van beleid X was een toename/afname van Y’ of ‘beleid X heeft geresulteerd in een toename/afname van Y’ zijn niet ongewoon.

Hier kleven echter allerlei problemen aan en deze worden ook in de beleidspraktijk herkend.^{xiii} Om te beginnen, het leert ons niet of het beleid echt werkt. Het enige dat mogelijk kan worden vastgesteld is een correlatie tussen het beleid en een verandering in beoogde condities. Of deze verandering causaal is toe te schrijven aan de beleidsimplementatie blijft in het ongewisse. Ten tweede, dit leert ons niets over hoe beleid werkt en dat zet een serieuze rem op het leren van evalueren. Sanderson verwoordt het als volgt: *klassieke goal-oriented evaluation is concerned more with establishing whether policies and programmes achieve their intended effects rather than with understanding how they achieve their effects (unintended as well as intended), and this has limited its contribution to policy learning.*^{xiv} Ten derde, het leidt onze aandacht af van ‘what works’.^{xv} Het vergelijken van de beginsituatie met de ex-post situatie sluit de ogen voor *onvoorziene opbrengsten* van beleid en vestigt de aandacht op wat gemeten kan worden in plaats van ook aandacht te hebben voor meer stilzwijgende effecten.^{xvi} Deze stilzwijgende neveneffecten kunnen zowel positief als negatief zijn. Tot slot, de veronderstelling dat alle condities minus de beleidsinterventie gelijk blijven is fundamenteel problematisch. Zo leert complexiteitswetenschap dat de sociale systemen waarin wordt geïntervenieerd veelzijdig en dynamisch zijn en dat beleidsinterventies hierin interacteren.^{xvii} Ook in de evaluatiepraktijk dringt dit besef door. Zo stelt het Nederlands Kennisinstituut voor Mobiliteitsbeleid: *het is lastig om een waargenomen ontwikkeling (...) te isoleren van het effect (...). Immers, vele factoren hebben tegelijkertijd aan die waargenomen ontwikkeling bijgedragen.*^{xviii} Het vaststellen van de netto-effecten van beleid is bijgevolg een fantasma, of zal althans slechts in zeer zeldzame gevallen mogelijk zijn. Het vaststellen van contributie van een beleidsinitiatief aan bepaalde beleidseffecten, eerder dan attributie van een beleidsinitiatief, zal vaak het maximaal mogelijke zijn. Om deze en meer redenen stellen evaluatie-deskundigen, zoals Michael Q. Patton, methoden voor die meer zijn gericht op leerprocessen tijdens beleidsimplementatie ‘in-situ’ hun situationele complexiteit.^{xix} Hoewel dergelijke pragmatische^{xx} evaluatiebenaderingen tekortkomingen van de gouden evaluatiestandaard lijken te tackelen, is het overdragen van geleerde in-situ lessen naar nieuwe beleidsinitiatieven problematisch. Over-contextualiseren beperkt generalisatie van evaluatie uitkomsten en daarmee de overdracht van beleidsleren. Wij zien in realistische evaluatie^{xxi} een evaluatiebenadering die in staat is de bovengenoemde problemen van klassieke evaluatieopvattingen te vermijden zonder daarbij te over-contextualiseren. In de volgende paragraaf zetten wij de beginselen van realistische evaluatie uiteen.

3. Tussen methode en praktijk: realistische evaluatie

Empirisch-technisch verloopt een realistische evaluatie op eenzelfde manier als klassiek sociaalwetenschappelijk onderzoek. Vanuit een theorie worden hypothesen opgesteld over het effect

van een beleidsinitiatief op een bepaalde uitkomst die vervolgens empirisch kunnen worden getest. Falsificatie kan dan helpen te bepalen of de hypothese klopt. Echter, specifiek voor realistische evaluatie is dat theorieën worden opgevat als ‘veranderingsmodellen’. Een essentieel onderdeel van een veranderingsmodel is verwachtingen over ‘hoe’ het effect tot stand komt. Veranderingsmodellen vinden we terug in beleid; in realistische evaluatie wordt beleid opgevat als een theorie van verandering. Een voorbeeld is de Nederlandse wijkenaanpak.^{xxii} Het doel daarvan is om van achterstandswijken weer prachtwijken te maken. Dit beleid bevat allerlei ideeën over ‘hoe’ dat moet worden gerealiseerd, zoals in Rotterdam-Zuid het herstructureren van een aantal woongebieden. Voor Vlaanderen kunnen we een voorbeeld uit het Decreet betreffende het beleid inzake de convenants gemeentelijke ontwikkelingssamenwerking aanhalen. Eén van de hoofddoelstellingen van de convenants is de gemeenten aan te moedigen een volwaardige speler in ontwikkelingssamenwerking te worden. De instapcriteria voor een convenant moeten helpen deze doelstelling te bewerkstelligen, onder meer: het hebben van een ambtenaar en schep en ontwikkelingssamenwerking, een erkende adviesraad, en de verbintenis tot één voltijds equivalent ambtenaar ontwikkelingssamenwerking drie jaar na sluiten convenant directe samenwerking.^{xxiii}

Beleidstheorieën kunnen afkomstig zijn van wetenschappelijke theorieën, maar evenzeer resulteren uit gezond verstand, resultaten van voorgaande evaluaties, ervaringen, enzovoort. Hoewel veranderingsmodellen inherent onderdeel zijn van beleid, worden ze vaak niet als dusdanig meegenomen bij beleidsevaluatie. Daar ligt de focus vooral op het enkel vergelijken van de situatie met en zonder het beleid, zonder expliciet mee te nemen hoe de verandering tot stand kwam.^{xxiv} Analytisch kunnen veranderingsmodellen worden beschouwd als zogenaamde CMO-configuraties. Realistische evaluatoren zijn gericht op het blootleggen van welke mechanismen (M) door een welbepaalde beleidsinterventie op gang worden gebracht, en in welke context (C) deze een welbepaalde beleidsuitkomst (O) tot stand brengen. De configuratie van deze drie componenten C, M en O is cruciaal. In een realistische evaluatie komt het er op neer de veronderstelde CMO-configuratie te expliciteren, en te testen.^{xxv}

De stelling dat ‘een beleidsinterventie werkt’, is volgens Pawson en Tilley, de grondleggers van de RE benadering, misleidend.^{xxvi} Een programma of beleidsinterventie zelf zal nooit rechtstreeks werken; er is nooit een rechtstreeks causaal verband tussen een beleidsinterventie en een beleidsuitkomst.^{xxvii} Het zijn eerder de hulpbronnen die worden verschaft aan bepaalde doelgroepen die een bepaalde gedragsverandering zullen uitlokken. Zo is de veronderstelling in het voorbeeld van de wijkenaanpak dat de herstructurering van woongebieden ervoor zorgt dat ze aantrekkelijker worden voor hogeropgeleiden met hogere inkomens, en dat dit de leefbaarheid van de wijken zal verbeteren. De verplichting tot aanstelling van één voltijds equivalent ambtenaar op het einde van de convenantsperiode ontwikkelingssamenwerking stoelt bijvoorbeeld op de veronderstelling dat versnippering in verantwoordelijkheid over de uitvoering van de convenant over meerdere personen het belang besteed aan ontwikkelingssamenwerking binnen de gemeente zou kunnen verminderen.^{xxviii}

Het causale verband zal steeds werken via het wijzigen van bepaalde mechanismen, in een bepaalde context. Mechanismen betreffen de eigenlijke logica achter bepaalde beleidsinterventies. Ze beschrijven *what [it] is about a program which makes it work*.^{xxix} Ze verwijzen naar de onderliggende processen of naar bepaalde attitudes die maken dat bepaalde actoren – zoals personen, groepen en organisaties – zullen reageren op de inzet van een overheidsinstrument.^{xxx} Denk bijvoorbeeld aan verkeersboetes die gebaseerd zijn op de veronderstelling dat mensen hun rijgedrag zullen aanpassen uit ‘angst’ voor een financiële sanctie. Deze ‘angst’ is het causale mechanisme. Het verklaart hoe een bepaalde beleidsmaatregel werkt, door actief het verband te leggen tussen een bepaalde beleidsmaatregel en een bepaalde beleidsuitkomst.^{xxxi} Hoewel dikwijls impliciet verondersteld als onderdeel van zogenaamde beleidstheorieën van ambtenaren of besluitvormers zijn deze mechanismen veelal niet uitdrukkelijk omschreven in beleidsprogramma’s, laat staan in evaluaties ervan. Het is dan

ook aan evaluatoren om ze te expliciteren, in samenwerking met de betrokkenen. Een bijkomende uitdaging is evenwel dat deze mechanismen zeer zelden onmiddellijk zintuigelijk waarneembaar zijn.

Mogelijk en waarschijnlijk werkt een mechanisme echter slechts in bepaalde contexten – denk aan een bepaalde sociaaleconomische klasse in het voorbeeld van de wijkenaanpak of personen met boete-aversief gedrag in het voorbeeld over verkeersboetes. Of in het geval van de convenants ontwikkelingssamenwerking toonde een evaluatie bijvoorbeeld aan dat het instrument het meest geschikt is op leest van middelgrote gemeenten, omwille van de planlast en –financiële verplichtingen.^{xxxii} Contextfactoren kunnen velerlei zijn, en verwijzen naar karaktereigenschappen van doelgroepen, fysieke factoren, maar ook technologische of economische omstandigheden.^{xxxiii} Ze kunnen zowel betrekking hebben op micro-, meso-, als macroniveau. Realistische evaluatie benadrukt dan ook de gelaagdheid en verwevenheid van het ‘systeem’ waarin de interventie plaatsvindt. Ter illustratie: wil een tewerkstellingsprogramma vruchten afwerpen dan is dit niet alleen afhankelijk van het profiel van de sollicitanten, maar ook van het daadwerkelijke banenaanbod in de markt. Realistische evaluatie integreert deze contextuele condities inherent in haar benadering. De recente realistische evaluatie van Leemans waarin onder meer de effectiviteit van het scholenbouwbeleid in Vlaanderen werd nagegaan is een ‘schoolvoorbeeld’ van het meenemen van contextfactoren van verschillende niveaus. Naast lokale actoren die betrokken zijn bij het bouwen, gebruiken en beheren van een schoolgebouw werd ook de ruimere context van scholengemeenschappen en lokale besturen, belangengroepen en de werking van de overheid in de evaluatie meegenomen.^{xxxiv} Het belang van contextuele condities maakt ook dat eenzelfde beleidsinitiatief, afhankelijk van de context waarin het wordt afgezet, verschillende effecten of uitkomsten kan ressorteren. Elk beleid kan dus gemengde uitkomsten patronen genereren, bedoeld of onbedoeld.^{xxxv} Of beleidsprogramma’s werken, dient bijgevolg te worden afgemeten aan al deze uitkomsten. Schematisch kunnen we de realistische benadering dan voorstellen zoals afgebeeld in figuur 2.^{xxxvi}

(figuur 2: realistische evaluatie)

4. Realistische evaluatie en klassieke evaluatie vergeleken

De verschillen met de klassieke benadering zijn duidelijk. Waar men klassiek de vraag centraal zal stellen ‘werkt het?’ kan de kern van realistische evaluatie worden samengevat met de slagzin ‘wat werkt voor wie in welke omstandigheden?’^{xxxvii} Door het erkennen van beleidsinterventies als CMO configuraties, kunnen belangrijke tekortkomingen van de klassieke benadering worden ondervangen. Vooreerst, het ontrafelen van de werking van mechanismen onderliggend aan bepaalde beleidsveranderingen kan een antwoord bieden op de vraag hoe beleid werkt en of het daadwerkelijk het beleid is dat een effect veroorzaakt. Daarnaast laat realistische evaluatie meer ruimte voor variatie in uitkomsten over verschillende contexten heen, ongeacht of die effecten gewenst of ongewenst zijn, bedoeld of onbedoeld.^{xxxviii} Het universeel slagen of falen van een beleidsprogramma is dus weinig plausibel, maar zal worden gedetermineerd door de contexten waarin het wordt geïmplementeerd. Verder leert realistische evaluatie dat het zoeken naar netto-effecten van beleid zinloos is. Het beleid isoleren van de ruimere context is niet wenselijk.^{xxxix} Eenzelfde beleidsinitiatief kan immers afhankelijk van de ruimere contextconstellatie totaal verschillende effecten genereren. En eenzelfde beleidsuitkomst kan door verschillende CMO-configuraties worden gegenereerd. De verschillen tussen RE en klassieke evaluatie zijn weergegeven in tabel 1, maar dit laat nog onverlet de vraag hoe evaluatie uitkomsten systematisch kunnen worden overgedragen naar andere beleidscontexten.

(tabel 1: verschillen klassieke en realistische evaluatie)

5. Leren van evaluaties

In dit ‘evidence-based’ tijdperk, worden evaluaties meer dan ooit opgezet om lessen te trekken uit uitgevoerd beleid met het idee dat dit toekomstig beleid kan informeren. Echter, indien we het belang van contextfactoren erkennen op de werking van overheidsinterventies dan roept dit de vraag op hoe we dan toch lessen kunnen trekken over verschillende contexten heen. Qualitative Comparative Analysis (QCA)^{xli} dient zich hier aan als veelbelovende methodiek.^{xlii} QCA is een methode die ervan uitgaat dat configuraties – d.w.z. combinaties van condities – voldoende en/of noodzakelijk zijn voor een bepaalde uitkomst. Dit betekent dat condities steeds worden beschouwd in de context van andere condities; iedere casus wordt opgevat als een configuratie. QCA houdt dus een holistische benadering in van de realiteit. Door het systematisch vergelijken van de configuraties kunnen patronen worden ontdekt (zie onderstaand), en minder relevante causale condities worden geïdentificeerd. Zonder in te gaan op de methodisch-technische details in het bestek van dit artikel, willen we de parallellen van QCA met de principes van realistische evaluatie onderstrepen.^{xliii} Om te beginnen, ook in QCA is het effect van de aan- of afwezigheid van een bepaalde factor (of constellatie van factoren) afhankelijk van de bredere context. Ten tweede, een combinatie van factoren zal cruciaal zijn om een bepaalde uitkomst te verklaren, eerder dan één enkele factor. Ten derde, verschillende combinaties kunnen leiden tot dezelfde uitkomsten. Wat QCA toevoegt aan RE is dat het in staat is deze complexe causaliteit in zekere mate te generaliseren om de overdracht van beleidslessen systematisch te faciliteren. QCA kan worden gebruikt voor een systematische paarsgewijze vergelijking van configuraties van mechanismen, contextelementen en uitkomsten. Een vergelijking kan wijzen op de noodzakelijke en/of voldoende (combinaties van) elementen die kunnen leiden tot een positieve (of een negatieve beleidsuitkomst). Toegepast op het voorbeeld van de wijkenaanpak, bijvoorbeeld: als in twee achterstandswijken de aanpak succes afwerpt, en ze verschillen in slechts één relevant contextelement, bijvoorbeeld grootte van de wijk, dan kan worden beargumenteerd dat dit element niet verklaart waarom het beleid werkt: of wijken nu groot of klein zijn, het beleid werkt. Een iteratieve en systematische herhaling van deze paarsgewijze vergelijking in alle 40 achterstandswijken, kan ons leiden tot verschillende constellaties van minimale configuraties die succesvol zijn, of falen. In toekomstige situaties kan kennis van deze minimale configuraties (patronen) worden gebruikt voor het ontwerpen van beleid, wat dus gebaseerd is op lessen uit het verleden.

6. Conclusie en vooruitzicht

In het voorgaande hebben we bedoeld een lans te breken voor realistische evaluatie als een belangrijke stap voorwaarts in het tot elkaar brengen van methode en praktijk in beleidsevaluaties. Deze potentie van RE wordt gelukkig ook in de beleidspraktijk in toenemende mate herkend.^{xliiii} Een uitdaging stelt zich vaak nog in het trekken van lessen uit deze realistische evaluaties. In deze bijdrage stelden we daarom kort het potentieel van QCA als een veelbelovende ontwikkeling die niet onterecht wordt gezien als alternatieve methodiek in beleidsevaluatie.^{xliiii} Vooral voor realistische evaluatie lijkt dit een logische methodiek die recht doet aan de vele contextfactoren die beleidsuitkomsten kunnen beïnvloeden *en* het tegelijkertijd toestaat om duidelijk communiceerbare patronen te genereren naar beleidsmakers toe. In QCA termen spreekt men in dit verband soms wel eens over causale ‘recepten’ voor een bepaalde uitkomst.^{xliiii} Dit laatste maakt dat QCA ook aantrekkelijk is vanuit politiek en beleidsmatig opzicht. In de introductie op ons betoog wezen wij al op de frustratie van politici en andere opdrachtgevers dat evaluaties veelal *niet in staat zijn om ondubbelzinnige antwoorden te geven op de vraag waarom programma’s falen of succesvol zijn.*^{xliiii} QCA kan een brug slaan tussen de wens van ondubbelzinnigheid en realistische evaluatie. Wij verwachten dan ook meer toepassingen in dit verband.

References

- i Leeuw, F. L. (2009). Evaluation policy in the Netherlands. *New Directions For Evaluation*, 2009(123), pp. 87-102.
- ii Varone, F., Jacob, S. & De Winter, L. (2005). Polity, politics and policy evaluation in Belgium. *Evaluation*, 11(3), pp. 253-273.
- iii Van Gils, G. H. C., Leeuw, F. L., Gildemyn, M. & Inberg, L. (2010). *Leren van evalueren: Onderzoek naar het gebruik van evaluatieonderzoek bij het ministerie van buitenlandse zaken*. Den Haag: Ministerie van Buitenlandse Zaken.
- Nijland, H., Wortelboer-van Donselaar, P. M., Korteweg, J. A. C. & Snellen, D. (2010). *Met de kennis van nu: Leren van evalueren: Een casestudy: A5 Verlengde Westrandweg*. Den Haag & Bilthoven: Planbureau voor de Leefomgeving & Kennisinstituut voor Mobiliteitsbeleid.
- iv Tieleman, B. & Van Humbeeck, P. (2007). *Evolutie in de evaluatie? Analyse van de Vlaamse milieubeleidsevaluatiepraktijk. MIRA-BE*. Aalst: Vlaamse Milieumaatschappij.
- v Fobé, E. & Brans, M. (2012). *Toekomstverkennen bij de Vlaamse overheid: Analyse van praktijken en doorwerking*. Leuven: Steunpunt Bestuurlijke Organisatie Vlaanderen.
- vi Kerseboom, J. (2008). Leren van evalueren: Over het niet-gebruik van beleidsinformatie. *Bestuurskunde*, 17(2), pp. 41-46.
- vii Kerseboom (2008, p. 43).
- viii Stern, E., Stame, N., Mayne, J., Forss, K., Davies, R. & Befani, B. (2012). *Broadening the range of designs and methods for impact evaluations*. London: Department for International Development.
- ix Abma, T. A. (1996). *Responsief evalueren: Discoursen, controversen en allianties in het postmoderne*. Delft: Eburon. Citaat: p. 7.
- x Over deze drie varianten van evaluatie, zie bijvoorbeeld: Howlett, M. & Ramesh, M. (1995). *Studying public policy: Policy cycles and policy subsystems*. Oxford: Oxford University Press. Zie pp. 170-175.
- xi Pawson, R. & Tilley, N. (1997). *Realistic evaluation*. London: Sage.
- xii De Peuter, B., De Smedt, J. & Bouckaert, G. (2007). *Handleiding beleidsevaluatie. Deel 1: Evaluatiedesign en -management*. Leuven: Steunpunt Bestuurlijke Organisatie Vlaanderen. Figuur: p. 147.
- xiii Bakker, R. (2012). *Evaluatie als fundering voor beleid*. Den Haag: Ministerie van Binnenlandse Zaken en Koninkrijksrelaties.
- xiv Sanderson, I. (2000). Evaluation in complex policy systems. *Evaluation*, 6(4), pp. 433-454. Citaat: p. 440.
- xv Cf. Kerseboom (2008).
- xvi Van Twist, M. J. W. & Verheul, W. J. (2010). Onvoorziene opbrengsten: Meer dan de tragiek van goede bedoelingen. *Beleid En Maatschappij*, 37(4), pp. 308-318.
- xvii Teisman, G. R., Van Buuren, M. W. & Gerrits, L. M. (Eds.). (2009). *Managing complex governance systems: Dynamics, self-organization and coevolution in public investments*. New York: Routledge.
- Byrne, D. S. (2011). *Applying social science: The role of social research in politics, policy and practice*. Bristol: The Policy Press.
- xviii Berveling, J., Groot, W., Lijesen, M., Savelberg, F. & Van der Werff, E. (2009). *Na het knippen van het lint: Het ex-post evalueren van infrastructuur*. Den Haag: Ministerie van Verkeer en Waterstaat. Citaat: p. 31.
- xix Patton, M. Q. (2011). *Developmental evaluation: Applying complexity concepts to enhance innovation and use*. New York: The Guilford Press.

-
- xx Pawson & Tilley (1997).
- xxi Pawson, R. (2002). Evidence-based policy: The promise of 'realist synthesis'. *Evaluation*, 8(3), pp. 340-358.
- Pawson & Tilley (1997).
- xxii Cf. Van Twist & Verheul (2010).
- xxiii De Bruyn, T., Huyse, H., De Peuter, B., Pattyn, V., & Brans, M. (2008). *Eerste evaluatie van het decreet betreffende het beleid inzake convenants gemeentelijke ontwikkelingssamenwerking uitgevoerd in opdracht van het Vlaams Agentschap voor Internationale Samenwerking*. K.U.Leuven: Hoger Instituut voor de Arbeid/Instituut voor de Overheid.
- De Peuter, B., & Pattyn V. (2010). Evaluatie van interbestuurlijk beleid via de beleidstheoriegedreven benadering. *Vlaams Tijdschrift voor Overheidsmanagement*, 15(3), pp. 18-28.
- xxiv Van Twist & Verheul (2010).
- xxv Pawson & Tilley (1997).
- xxvi Pawson & Tilley (1997, p. 6).
- xxvii Pawson, R. & Tilley, N. (2004). *Realist evaluation*. London: British Cabinet Office.
- xxviii De Bruyn, T., Huyse, H., De Peuter, B., Pattyn, V., & Brans, M. (2008).
- xxix Pawson & Tilley (1997, p. 66).
- xxx Leeuw, F. L. & Vaessen, J. (2010). Introduction: Mind the gap. In J. Vaessen & F. L. Leeuw (Eds.), *Mind the gap: Perspectives on policy evaluation and the social sciences* (pp. 1-29). London: Transaction Publishers.
- xxxi Leemans, G. (2013). *Een realistische evaluatie van het scholenbouwbeleid in Vlaanderen: Logica en coherentie, effectiviteit, actiepunten*. Brussel: Agentschap voor Infrastructuur in het Onderwijs.
- xxxii De Bruyn, T., Huyse, H., De Peuter, B., Pattyn, V., & Brans, M. (2008, p. 70).
- xxxiii Pawson & Tilley (2004) en Bakker (2012).
- xxxiv Leemans (2013).
- xxxv Zie bijvoorbeeld: Van Twist & Verheul (2010).
- xxxvi Pawson & Tilley (1997, p. 58).
- xxxvii Pawson & Tilley (1997, p. 210).
- xxxviii Van Twist & Verheul (2010) en Leemans (2013).
- xxxix Verweij, S. & Gerrits, L. M. (2013). Understanding and researching complexity with qualitative comparative analysis: Evaluating transportation infrastructure projects. *Evaluation*, 19(1), pp. 40-55.
- xl Ragin, C. C. (1987). *The comparative method: Moving beyond qualitative and quantitative strategies*. Los Angeles: University of California Press.
- xli Verweij, S. & Gerrits, L. M. (2012). Systematische kwalitatieve vergelijkende analyse. *Kwalon: Tijdschrift Voor Kwalitatief Onderzoek*, 17(3), pp. 25-33.
- xlii Befani, B., Ledermann, S. & Sager, F. (2007). Realistic evaluation and QCA: Conceptual parallels and an empirical application. *Evaluation*, 13(2), pp. 171-192.
- Befani, B. & Sager, F. (2006). QCA as a tool for realistic evaluations: The case of the Swiss environmental impact assessment. In B. Rihoux & H. Grimm (Eds.), (pp. 263-284). New York: Springer.
- Sager, F. & Andereggen, C. (2012). Dealing with complex causality in realist synthesis: The promise of qualitative comparative analysis. *American Journal Of Evaluation*, 33(1), pp. 60-78.
- xliii Zie bijvoorbeeld: Van Gils et al. (2010), Bakker (2012), Stern et al. (2012) en Leemans (2013).

-
- xliv Marx, A. (2005). Systematisch comparatief case onderzoek van evaluatieonderzoek. *Tijdschrift Voor Sociologie*, 26(1), pp. 95-113.
- Varone, F., Rihoux, B. & Marx, A. (2006). A new method for policy evaluation? Longstanding challenges and the possibilities of qualitative comparative analysis (QCA). In B. Rihoux & H. Grimm (Eds.), *Innovative comparative methods for policy analysis: Beyond the quantitative-qualitative divide* (pp. 213-236). New York: Springer.
- xlv Ragin, C. C. (2008). *Redesigning social inquiry: Fuzzy sets and beyond*. Chicago: University of Chicago Press.
- xlvi Abma (1996, p. 7).