

## ARTIKEL

# Hoe weet je wat werkt? Over de zoektocht naar bruikbare kennis voor beleidseffectiviteit

*Gerdien van Eersel\**

‘Intuïtie kan zo schadelijk zijn’, kopte *de Volkskrant*<sup>1</sup> vorig jaar. Als je volgens nobelprijswinnaar Esther Duflo beleid baseert op intuïtie, dan maak je beleid dat mogelijk niet werkt. Duflo wilde met haar onderzoek intuïties over het effect van onderwijsinterventies ontcrachten. Zij en haar collega’s lieten met gerandomiseerde onderzoeksmethoden zien dat de onderwijsprestaties in Kenia niet verbeterden door meer schoolboeken of een gratis maaltijd, maar door een medicijn tegen wormen waarmee de kinderen waren geïnfecteerd (Duflo & Kremer, 2003). Beleid zou daarom in het algemeen meer ‘evidence-based’ of ‘evidence-informed’ moeten zijn (voor het onderscheid, zie o.a. Nevo & Slonim-Nevo, 2011), dat wil zeggen, onderbouwd met wetenschappelijk bewijs. Maar welk wetenschappelijk bewijs is eigenlijk relevant voor beleid? Oftewel, hoe weet je wat werkt?

Om te voorspellen of met een beleidsinterventie de gewenste uitkomsten worden gerealiseerd, verwijst onder andere het ministerie van Financiën naar een van de bekende hiërarchieën van bewijs, namelijk de Maryland Scientific Methods Scale (MSMS).<sup>2</sup> Met de MSMS kan de kwaliteit van een onderzoek worden bepaald aan de hand van een aantal gradaties van interne validiteit. De interne validiteit van een onderzoek betreft de mogelijkheid van het vaststellen van de causale relatie tussen interventie en effect: als we X doen, dan bereiken we Y. De Randomized Controlled Trial (RCT) geldt hier als de gouden standaard; deze onderzoeksmethode levert bewijs van de hoogste kwaliteit.

Het design van de RCT maakt het mogelijk om te bepalen of een beleidsinterventie een causale rol speelt voor de voorspelde uitkomst. Met een RCT kan worden aangetoond dat een interventie een effect had op de uitkomstmaat, dat wil zeggen dat het beleid *ergens* werkte, binnen een bepaalde context. Dit is echter niet voldoende; het is noodzakelijk om te weten of de beleidsinterventie het effect zal hebben binnen de context waarin de interventie wordt geïmplementeerd (Cartwright & Munro, 2010). In dat geval is de interventie ook *effectief*. Deze effectiviteit is niet zomaar vanzelfsprekend (zie o.a. Onrust et al., 2016; Sundell et al., 2016). Bijvoorbeeld, een naschools programma om tienerzwangerschap te voorkomen bleek ef-

\* Gerdien van Eersel is universitair docent Interdisciplinaire Sociale Wetenschap aan de Universiteit Utrecht. Ze was daarvoor onderzoeker bij de lokale overheid en heeft een achtergrond in de wetenschapsfilosofie.

1 Zie [www.volkskrant.nl/economie/nobelprijswinnaar-esther-duflo-intuïtie-kan-zo-schadelijk-zijn~b4284201/](http://www.volkskrant.nl/economie/nobelprijswinnaar-esther-duflo-intuïtie-kan-zo-schadelijk-zijn~b4284201/).

2 Zie [www.rijksbegroting.nl/beleidsevaluaties/evaluaties-en-beleidsdoorlichtingen/handreiking-beleidsdoorlichtingen/het-meten-van-doeltreffendheid](http://www.rijksbegroting.nl/beleidsevaluaties/evaluaties-en-beleidsdoorlichtingen/handreiking-beleidsdoorlichtingen/het-meten-van-doeltreffendheid).

Gerdien van Eersel

fectief in New York maar niet in andere delen van de VS of in Engeland (Bonell et al., 2012). Deze verschillen kunnen deels worden verklaard aan de hand van contextverschillen. Zo werd het programma in Engeland ingezet als alternatief voor normaal onderwijs, waardoor deelnemers regulier onderwijs moesten missen. Verder kwamen in New York alle jongeren binnen een bepaald gebied in aanmerking voor de interventie, terwijl in Engeland de interventie gericht was op specifieke kwetsbare jongeren. Dit kan nadelige bijeffecten hebben gehad via negatieve peer-invloeden in de sociale netwerken van de jongeren.

Welk type kennis is nu nodig om te kunnen inschatten of een interventie effectief zal zijn? Op basis van welke inzichten kunnen we verwachten dat een beleidsmaatregel in de praktijk zal uitwerken zoals bedoeld? Dit vereist meer kennis dan een RCT of meta-analyse kan leveren (Pawson, 2002a), namelijk kennis over hoe beleid binnen een specifieke context zal uitwerken. Twee typen kennisclaims zijn hiervoor specifiek van belang, namelijk die over de mechanismes tussen interventie en effect (Marchioni & Reijula, 2019) en die over de benodigde ondersteunende factoren (Cartwright & Hardie, 2012). Dit zal worden toegelicht aan de hand van twee voorbeelden. NB, dit artikel bevat geen nieuwe visies op beleidsevaluatie, maar bespreekt welke kennis relevant is voor het ontwikkelen en uitvoeren van beleid. Deze informatie is daarmee ook van belang voor *ex ante*, *ex durante* en *ex post* beleidsevaluatie.

## Verklarend mechanisme

Een mechanisme geeft een theoretische onderbouwing van hoe en waarom een (beleids)maatregel een zekere uitkomst zal hebben. Een dergelijk ‘verklarend’ of ‘causaal’ mechanisme staat gelijk aan een proces dat een aantal opeenvolgende gebeurtenissen of variabelen causaal met elkaar verbindt (Woolcock, 2019). Een mechanisme komt binnen een bepaalde context tot stand als gevolg van een beleidsmaatregel, en heeft een zekere uitkomst tot gevolg (Van der Knaap et al., 2020: 100). Bijvoorbeeld, Donohue en Levitt (2020) laten zien dat de legalisering van abortus zorgde voor een afname van het aantal misdrijven twintig jaar later. De verklaring hiervoor is dat de geboortedaling als gevolg van de legalisering het hoogst was onder jonge, vaak alleenstaande vrouwen zonder opleiding (Donohue & Levitt, 2001). Omdat deze familieomstandigheden samenhangen met delinquentie op jongere leeftijd (Donohue & Levitt, 2001), zorgde de legalisering van abortus zo voor een afname van de criminaliteit twintig jaar later.

Recent zijn er ook verschillende context-mechanisme-uitkomst-evaluaties verschenen, waarbij de contextuele factoren alsook de mechanismes van een maatregel in kaart worden gebracht (Bouman-van 't Veer et al., 2011; De Brún & McAuliffe, 2020; De Souza, 2013; Herens et al., 2017). Zo laten Rycroft-Malone et al. (2010) zien dat de invoering van gestandaardiseerde zorg in Engeland leidde tot meer vrijheid en autonomie van zorgprofessionals vanwege de gestandaardiseerde protocollen. Hierdoor verruimde het takenpakket van het verplegend en verloskundig personeel en konden zij taken overnemen van artsen, die daarmee weer werden ontlast. De relevante context was in dit geval het type kliniek waar het verplegend personeel al veelal zelfstandig kon werken (zoals geboortecentra en kli-

nieken voor chronische ziekten), waardoor het betreffende mechanisme kon worden geactiveerd.

Een ander voorbeeld van een mechanisme tussen beleidsmaatregel en uitkomst wordt gegeven door Pawson (2002b), die de veronderstelde werking van een verplicht notificatiesysteem van de adressen van zedendelinquenten ('Megan's Law' uit 1996) heeft gereconstrueerd. Het doel van deze nieuwe wet was het terugdringen van recidive. Pawson (2002b; 2003) heeft in een literatuurstudie uiteengezet welk bewijs er was voor de opeenvolgende componenten van het mechanisme. Het mechanisme bestaat uit vier componenten. De eerste component is het identificeren en betrouwbaar registreren van veroordeelde zedenplegers met een hoog risico op recidive. De tweede stap is het publiek bekendmaken van deze geregistreerde delinquenten. De derde stap is dat de geïnformeerde gemeenschap zich aansluit bij de politie, zodat een surveillancesysteem ontstaat waarmee verdacht gedrag eerder wordt opgemerkt. De vierde stap is de reactie van de delinquent, die door het toezicht en de stigmatisering minder snel in herhaling zal vallen.

Pawson (2002b; 2003) bespreekt in zijn literatuurstudie een quasi-experimenteel onderzoek (Schram & Milloy, 1995) dat de vierde stap van het mechanisme onderzocht: was er minder recidive met het notificatiesysteem dan zonder het notificatiesysteem? Uit de resultaten bleek dat dit nauwelijks het geval was. Daarnaast bespreekt Pawson (2002b; 2003) een simulatiestudie (Petrosino & Petrosino, 1999) met als onderzoeksvraag de eerste en derde stap van het mechanisme: in welk percentage van de zedendelicten zou het notificatiesysteem de kans hebben verhoogd dat verdacht gedrag eerder was opgemerkt? Dit percentage bleek heel klein, omdat van alle bestudeerde delinquenten maar een klein deel aan de juiste criteria voldeed. Ook bleek het lastig om het recidive-risico goed te voorspellen. Als laatste bespreekt Pawson een kwalitatief onderzoek onder reclasseringsambtenaren (Zevitz & Farkas, 2000a). Hierin werd de aanname van de derde stap van het mechanisme uitgevraagd: nam de samenwerking tussen gemeenschap en politie toe en werd hierdoor de surveillance effectiever? De antwoorden van de reclasseringsambtenaren gaven een ander beeld. Zij waren vooral bezig met het beschermen van delinquenten tegen intimidatie vanuit de gemeenschap, die niet zozeer samenwerkte met de politie maar vooral zelf initiatief nam. In een ander kwalitatief onderzoek (Zevitz & Farkas, 2000b) gaven zedendelinquenten nog aan dat zij bang waren voor repercussies van buurtgenoten als die op de hoogte waren van hun criminele verleden, wat maakte dat zij de registratie probeerden te omzeilen. Een van de noodzakelijke ondersteunende factoren voor een succesvolle implementatie van Megan's Law was echter juist een betrouwbaar registratiesysteem. Dit alles kan gevolgen hebben gehad voor de werking en het succes van het systeem.

Gezamenlijk laten deze onderzoeken zien dat Megan's Law niet werkte zoals bedoeld en wat daarvoor de redenen waren. Verschillende typen onderzoeken maken deel uit van Pawsons literatuurstudie. Slechts een beperkt deel van de besproken onderzoeken betrof een academische studie, en niet alle onderzoeken kunnen direct worden geassocieerd volgens de MSMS. Een deel van de onderzoeken was tevens kwalitatief van aard, wat een geschikte methode is om kennis te vergaren over mechanismes en contextuele factoren (o.a. Woolcock, 2019).

Gerdien van Eersel

Meer algemeen laat deze studie zien dat de relevantie van een onderzoek niet kan worden bepaald op basis van een algemeen geldende bewijshiërarchie zoals de MSMS. Het enige criterium om vast te stellen of een onderzoek inzicht geeft in de effectiviteit van beleid, is door te bepalen of het onderzoek het begrip vergroot van het onderliggende mechanisme of theorie (Watts et al., 2011). Op deze manier dragen ook de besproken onderzoeken bij aan de kennis over het mechanisme rondom de werking van Megan's Law. Het identificeren van het onderliggende mechanisme versterkt bovendien de kennis over de generaliseerbaarheid van effecten (Cinelli & Pearl, 2020) en de benodigde ondersteunende factoren (Cartwright & Stegenga, 2011). Wat die ondersteunende factoren ofwel randvoorwaarden precies inhouden, wordt verder geïllustreerd in het volgende voorbeeld.

### Ondersteunende factoren

De inzet van kleinere schoolklassen is een langgekoesterde wens van onderwijsorganisaties, en staat sinds de eerste lockdown in 2020 weer op de agenda.<sup>3</sup> Het zou een middel zijn om de werkdruk te verlagen en om schoolprestaties te verbeteren. Maar werkt het ook? In Tennessee is al in 1985 onderzocht of de schoolprestaties van basisschoolleerlingen verbeterden in kleine klassen (13-17 leerlingen) vergeleken met reguliere klassen (22-25 leerlingen). Dit onderzoek werd uitgevoerd met een RCT onder 6572 leerlingen, die willekeurig werden toegewezen aan een van de twee klasgroottes. Vier jaar later bleek dat de leerlingen uit de kleine klassen beter waren geworden in lezen en schrijven dan de leerlingen uit de reguliere klassen (Mosteller, 1995). Deze verschillen waren het grootst bij leerlingen uit gezinnen met lagere inkomens of uit etnische minderheidsgroepen (Bohrnstedt & Stecher, 2002; Mosteller, 1995).

Omdat de staat Californië kampte met lage schoolprestaties, werd in 1996 op basis van het bewijs uit Tennessee besloten om de interventie van kleinere klassen in heel Californië in te voeren. In 2002 werd hierover een onderzoeksrapport gepubliceerd met de opbrengsten van de maatregel (Bohrnstedt & Stecher, 2002). Er bleek geen bewijs te zijn dat de inzet van kleinere klassen in Californië leerwinst had opgeleverd.

Dit voorbeeld illustreert dat een RCT slechts laat zien dat een interventie werkt binnen een bepaalde setting, in dit geval op geselecteerde scholen in Tennessee. De interventie werkte niet in Californië. Wat verklaart het verschil tussen de twee staten? Allereerst werd in Californië het programma in slechts een jaar uitgerold, waardoor er een tekort aan klaslokalen ontstond (Bohrnstedt & Stecher, 2002). In Tennessee werden alleen de scholen bij het programma betrokken die genoeg beschikbare ruimte hadden. Ook werden in Californië lokalen gebruikt die normaal ingezet werden voor activiteiten zoals sport en kinderopvang, en moest er worden bezuinigd op bijvoorbeeld onderhoud en software vanwege de hoge kosten van de kleine klassen (Bohrnstedt & Stecher, 2002). Daarnaast was Californië genoodzaakt om 12.000 extra leerkrachten in te huren, van wie een groot deel ongekwalifi-

3 Zie [www.trouw.nl/onderwijs/petitie-gestart-voor-kleinere-klassen~b96ab880/](http://www.trouw.nl/onderwijs/petitie-gestart-voor-kleinere-klassen~b96ab880/).

ficeerd. Veel van deze leerkrachten werden ingezet op scholen met een hoog percentage kwetsbare leerlingen (Bohrnstedt & Stecher, 2002).

Dit alles betekent dat in Californië niet aan alle noodzakelijke randvoorwaarden was voldaan om zo het prestatieverhogende effect van kleinere klassen mogelijk te maken. Kleinere klassen bleken op zichzelf niet voldoende om de prestaties te verhogen; de benodigde randvoorwaarden waren genoeg klaslokalen en gekwalificeerde leerkrachten (Cartwright & Hardie, 2012). Als één van die factoren niet voorhanden is, dan treedt het effect niet of in geringere mate op. Met andere woorden, de factoren vormen gezamenlijk het causale pakket dat het effect veroorzaakt (Cartwright & Hardie, 2012). Alleen als alle factoren tegelijk aanwezig zijn, dan zal de interventie het verwachte effect hebben.<sup>4</sup> Deze analyse correspondeert met de *Qualitative Comparative Analysis*, een evaluatiemethode met als doel om de noodzakelijke en voldoende voorwaarden vast te stellen die een bepaalde uitkomst verklaren (voor een overzicht, zie Stinckens, 2016). Hierbij kunnen dan verschillende causale paden naar de uitkomstvariabele worden geïdentificeerd.

Er bestaan ook nog andere interventies die bijdragen aan leerprestaties, zoals thuis voorlezen (Mol & Bus, 2011) of aanvullend onderwijs voor extra ondersteuning (Bolhaar et al., 2018). Ook voor deze interventies geldt dat het positieve effect ervan alleen optreedt wanneer ze worden gecombineerd met een aantal specifieke ondersteunende factoren (Cartwright & Hardie, 2012). Bijvoorbeeld, als gevolg van de kleinere klassen waren er in Californië minder lokalen beschikbaar voor aanvullend onderwijs aan kwetsbare kinderen (Bohrnstedt & Stecher, 2002). Een voorwaarde voor het succes van aanvullend onderwijs is echter juist de beschikbaarheid van voldoende klaslokalen. Omdat deze factor wegviel met het introduceren van de kleinere klassen, werd daarmee mogelijk ook het positieve effect van het aanvullend onderwijs verkleind (Cartwright & Hardie, 2012).

## Relevantie voor beleid

Welke kennisclaims zijn nodig om zicht te krijgen op de te verwachten effecten van beleid? Dit zijn claims die kennis opleveren over de mechanismes tussen interventie en effect en over de noodzakelijke randvoorwaarden. Dit standpunt bevindt zich op het raakvlak van de rationeel-technische en contextueel-realistische benadering van beleidsonderzoek. De eerstgenoemde betreft voornamelijk kwantitatief onderzoek en is sterk gericht op effectiviteit (Van der Knaap et al., 2020). De contextueel-realistische benadering gaat vooral over de vraag wat werkt binnen welke context en op welke manier (Pawson & Tilley, 1997). Daarnaast bestaat er nog het responsieve beleidsonderzoek, waarbij de ervaringen van betrokkenen met het beleid en de effecten ervan centraal staan.

De kracht van de rationeel-technische benadering is de nadruk op de vraag naar effectiviteit: worden als gevolg van het beleid de beoogde uitkomsten gerealiseerd? Een beleidsinterventie is effectief indien deze leidt tot de beoogde uitkomst. De contextueel-realistische benadering voegt hier het belang van context aan toe. Een

4 Dit sluit aan bij Mackies (1965) definitie van oorzaken als 'INUS'-condities: 'Insufficient but Necessary parts of a condition which is itself Unnecessary but Sufficient'.

Gerdien van Eersel

beleidsinterventie kan verschillende mechanismes activeren, afhankelijk van de context. Om de vraag te beantwoorden of het beleid zal leiden tot de gewenste uitkomsten, is het daarom noodzakelijk om de verklarende mechanismes en bijbehorende context te achterhalen. Binnen deze benadering gaat het dus om het formuleren en testen van deze context-mechanisme-uitkomst-configuraties (Pawson et al., 2005). Om inzicht te krijgen in de manier waarop beleid in de praktijk zal uitwerken, is het dus essentieel om te weten of een beleidsinterventie binnen een bepaalde context zal uitpakken zoals beoogd.

Bij het ontwikkelen van beleidsinterventies dient de kennis over verklarende mechanismes en ondersteunende factoren centraal te staan. Wat zijn de werkzame componenten van het verklarende mechanisme? Wat zijn de benodigde ondersteunende factoren, en welke daarvan zijn aanwezig in de betreffende context? Welke causale factoren zijn onderzocht, en welke bestaan binnen de context in kwestie? Op basis van deze kennis kan worden ingeschat of met de inzet van de interventie het gewenste doel zal worden bereikt. Deze voorspelling kent altijd een bepaalde mate van onzekerheid, en ook zal de lijst van causale factoren bijna nooit volledig zijn. De benadering is echter wel van grote waarde om scherper zicht te krijgen op de te verwachten effecten van beleid.

*Met dank aan Eric van Straten en Peter van der Knaap voor de feedback op eerdere versies van dit manuscript.*

## Literatuur

- Bohrnstedt, G., & Stecher, B. (2002). *What we have learned about class size reduction in California*. Sacramento: California Department of Education. Retrieved from <https://edsources.org/wp-content/publications/CSRSummaryFinal.pdf>
- Bolhaar, J., Houkes-Hommes, A., & Van der Ven, K. (2018). Bewezen (in)effectieve maatregelen tegen leerachterstanden in het primair onderwijs. *CPB Achtergronddocument*. CPB-Achtergronddocument-6jun2018-Bewezen-in-effectieve-maatregelen-tegen-leerachterstanden-in-het-primair-onderwijs.pdf
- Bonell, C., Fletcher, A., Morton, M., Lorenc, T., & Moore, L. (2012). Realist randomized controlled trials: A new approach to evaluating complex public health interventions. *Social Science & Medicine*, 75, 2299-2306. doi:10.1016/j.socscimed.2012.08.032
- Bouman-van 't Veer, M., Knijn, T., & Van Berkel, R. (2011). *Activeren door te participeren: De meerwaarde van de Wet maatschappelijke ondersteuning voor re-integratie van mensen in de bijstand*. Utrecht: Movisie.
- Cartwright, N., & Hardie, J. (2012). *Evidence-based policy: A practical guide to doing it better*. Oxford: Oxford University Press.
- Cartwright, N., & Munro, E. (2010). The limitations of randomized controlled trials in predicting effectiveness. *Journal of Evaluation in Clinical Practice*, 16(2), 260-266. doi:10.1111/j.1365-2753.2010.01382.x
- Cartwright, N., & Stegenga, J. (2011). A theory of evidence for evidence-based policy. In P. Dawid, W. Twining, & M. Vasilaki (Eds.), *Evidence, inference and enquiry* (pp. 291-322). Oxford: Oxford University Press.
- Cinelli, C., & Pearl, J. (2020). Generalizing experimental results by leveraging knowledge of mechanisms. *European Journal of Epidemiology*. doi:10.1007/s10654-020-00687-4

- De Brún, A., & McAuliffe, E. (2020). Identifying the context, mechanisms and outcomes underlying collective leadership in teams: Building a realist programme theory. *BMC Health Services Research*, 20, 261. <https://doi.org/10.1186/s12913-020-05129-1>
- De Souza, D.E. (2013). Elaborating the context-mechanism-outcome configuration (CMOC) in realist evaluation: A critical realist perspective. *Evaluation*, 19, 141-154. doi:<https://doi.org/10.1177/1356389013485194>
- Donohue, J.J., & Levitt, S.D. (2001). The impact of legalized abortion on crime. *The Quarterly Journal of Economics*, 116(2), 379-420.
- Donohue, J.J., & Levitt, S.D. (2020). The impact of legalized abortion on crime over the last two decades. *American Law and Economics Review*, 22, 241-302. doi:<https://doi.org/10.1093/aler/ahaa008>
- Duflo, E., & Kremer, M. (2003). *Use of randomization in the evaluation of development effectiveness*. Paper prepared for the World Bank Operations Evaluation Department (OED) Conference on Evaluation and Development Effectiveness in Washington, DC. <https://economics.mit.edu/files/765>
- Herens, M., Wagemakers, A., Vaandrager, L., Van Ophem, J., & Koelen, M. (2017). Contexts, mechanisms, and outcomes that matter in Dutch community-based physical activity programs targeting socially vulnerable groups. *Evaluation & Health Professions*, 40, 294-331. doi:<https://doi.org/10.1177/0163278716652940>
- Mackie, J.L. (1965). Causes and conditions. *American Philosophical Quarterly*, 12, 245-265.
- Marchioni, C., & Reijula, S. (2019). What is mechanistic evidence, and why do we need it for evidence-based policy? *Studies in History and Philosophy of Science part A*, 73, 54-63.
- Mol, S., & Bus, A. (2011). Lezen loont een leven lang: De rol van vrijetijdslezen in de taal- en leesontwikkeling van kinderen en jongeren. *Levende Talen Tijdschrift*, 12, 3-15.
- Mosteller, F. (1995). The Tennessee study of class size in the early school grades. *The Future of Children*, 5, 113-127.
- Nevo, I., & Slonim-Nevo, V. (2011). The myth of evidence-based practice: Towards evidence-informed practice. *The British Journal of Social Work*, 4, 1176-1197. <https://doi.org/10.1093/bjsw/bcq149>
- Onrust, S.A., Otten, R., Lammers, J., & Smit, F. (2016). School-based programmes to reduce and prevent substance use in different age groups: What works for whom? Systematic review and meta-regression analysis. *Clinical Psychology Review*, 44, 45-59. doi:10.1016/j.cpr.2015.11.002
- Pawson, R. (2002a). Evidence-based policy: In search of a method. *Evaluation*, 8, 157-181.
- Pawson, R. (2002b). *Does Megan's Law work? A theory-driven systematic review*. London: ESRC Centre for Evidence Based Policy and Practice. Retrieved from <https://www.kcl.ac.uk/sspp/departments/politiceconomy/research/cep/pubs/papers/assets/wp8.pdf>
- Pawson, R. (2003). Nothing as practical as a good theory. *Evaluation*, 9, 309-321.
- Pawson, R., Greenhalgh, T., Harvey, G., & Walshe, K. (2005). Realist review – a new method of systematic review designed for complex policy interventions. *Journal of Health Services Research & Policy*, 10, 21-34.
- Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.
- Petrosino, A., & Petrosino, C. (1999). The public safety potential of Megan's Law in Massachusetts: An assessment from a sample of criminal sexual psychopaths. *Crime and Delinquency*, 45, 140-158.
- Rycroft-Malone, J., Fontenla, M., Bick, D., & Seers, K. (2010). A realistic evaluation: The case of protocol-based care. *Implementation Science*, 5, 38. <https://doi.org/10.1186/1748-5908-5-38>
- Schram, D., & Milloy, C. (1995). *Community notification: A study of offender characteristics and recidivism*. Olympia: Washington State Institute for Public Policy. Retrieved from [http://www.wsipp.wa.gov/ReportFile/1208/Wsipp\\_Community-Notification-A-Study-of-Offender-Characteristics-and-Recidivism\\_Full-Report.pdf](http://www.wsipp.wa.gov/ReportFile/1208/Wsipp_Community-Notification-A-Study-of-Offender-Characteristics-and-Recidivism_Full-Report.pdf)

Gerdien van Eersel

- Stinckens, M. (2016). Op zoek naar innovatie in het beleidsimplementatieonderzoek: de meerwaarde van 'Qualitative Comparative Analysis'. *Beleidsonderzoek Online*. doi:10.5553/BO/221335502016000002001
- Sundell, K., Beelmann, A., Hasson, H., & Von Thiele Schwarz, U. (2016). Novel programs, international adoptions, or contextual adaptations? Meta-analytical results from German and Swedish intervention research. *Journal of Clinical Child & Adolescence Psychology*, 45, 784-796. doi:10.1080/15374416.2015.1020540
- Van der Knaap, P., Pattyn, V., & Hanemaayer, D. (2020). *Beleidsevaluatie in theorie en praktijk*. Den Haag: Boom bestuurskunde.
- Watts, P., Phillips, G., & Petticrew, M. (2011). The influence of environmental factors on the generalisability of public health research evidence: Physical activity as a worked example. *International Journal of Behavioral Nutrition and Physical Activity*, 8. doi:10.1186/1479-5868-8-128
- Woolcock, M. (2019). Reasons for using mixed methods in the evaluation of complex projects. In M. Nagatsu & A. Ruzzene (Eds.), *Contemporary philosophy and social science* (pp. 147-172). Bloomsbury Academic. doi:10.5040/9781474248785.ch-005
- Zevitz, R., & Farkas, M. (2000a). The impact of sex-offender community notification on probation/parole in Wisconsin. *International Journal of Offender Therapy and Comparative Criminology*, 44, 8-21.
- Zevitz, R., & Farkas, M. (2000b). Sex offender community notification: Managing high risk criminals or exacting further vengeance? *Behavioural Sciences and the Law*, 18, 375-392.