

Effectevaluatie wetenschappelijk

In 2004 publiceerden Wittebrood en Van Beem (SCP) in opdracht van de Raad voor Maatschappelijke Ontwikkeling (RMO) een achtergrondstudie naar de vraag 'wat wel en wat niet werkt om de sociale veiligheid in onze samenleving te vergroten'. Deze studie, die zich duidelijk positioneert in de 'What works'-traditie (Sherman, *Evidence-based crime prevention*, 2002), kwam tot de ontluisterende conclusie dat voor de talloze maatregelen die de overheid vanaf begin jaren tachtig heeft genomen op het gebied van sociale veiligheid, slechts tien effectevaluatie konden worden gevonden waarvan het onderzoeksdesign conclusies toeliet over de effectiviteit van de maatregel. Voor hun oordeel over de kwaliteit van de studies gebruikten Wittebrood en Van Beem criteria zoals opgesteld door de Campbell Collaboration. Deze organisatie, die als credo de leus voert: *What helps? What harms? Based on what evidence?*, houdt zich bezig met het faciliteren, uitvoeren en beschikbaar stellen van systematische overzichtsstudies naar effecten van sociale interventies in onderwijs, welzijn en criminaliteitspreventie en rechtshandhaving (zie: <http://www.campbellcollaboration.org>).

Hoewel de conclusie van Wittebrood en Van Beem niet exemplarisch is voor de staat van onderzoek naar sociale interventies in Nederland, zijn er enkele interessante voorbeelden aan te wijzen waarin wél sprake is van evaluatie die aan de eisen van de Campbell Collaboration voldoet. Voorbeelden op het gebied van criminaliteitspreventie zijn de evaluatie van de HALT-maatregel door Ferwerda et al. (WODC, 2006), het experiment naar effecten van de benadering van zwartrijders (Bijleveld in het *Journal of Experimental Criminology*, 2007, 3) en de effectevaluatie Strafrechtelijke Opvang Verslaafden (Koeter & Bakker, WODC, 2007). Het aantal van dergelijke studies is echter nog klein.

Verder valt op dat er een hardnekkige scheiding bestaat tussen de wereld van de 'correlationele', veelal op grootschalige surveyonderzoeken gebaseerde studies en die van de toegepaste effectstudies. Dat is jammer, want beide tradities zouden elkaar goed aan kunnen vullen. Daar waar auteurs van 'correlationele' studies zich al wagen aan uitspraken over effecten doen zij dat op basis van geavanceerdere analysemethoden die, gebruikmakend van non-experimentele dataverzamelingsdesigns, het experimentele design zoveel mogelijk proberen te benaderen (zie bijvoorbeeld het artikel van Nieuwbeerta et al. in het derde nummer van *Mens en Maatschappij* uit 2007). Het is echter vaak maar beperkt mogelijk conclusies te trekken over de effectiviteit van interventies aan de hand van dit type studies.

Ook aan de (quasi-)experimentele effectstudies kleven belangrijke bezwaren. Zo is het ontwerpen en beheersen van een experimenteel design en het interpreteren van de uitkomsten een complexe onderneming. In de eerste plaats is de sociale interventie zelf maar in beperkte mate te controleren (interne validiteitsprobleem). Het belangrijkste daarbij is misschien wel dat de interventie niet wordt uitgevoerd zoals bedoeld en dat mensen die toegewezen zijn aan

een interventie, meer dan bij bijvoorbeeld medisch onderzoek, uiteindelijk niet de (gehele) interventie ondergaan. In de tweede plaats is er het probleem met de externe validiteit van de conclusies. Interventies vinden altijd onder bepaalde omstandigheden plaats en juist die omstandigheden zijn een eventuele volgende keer weer anders. Het gebeurt dan ook vaak dat een interventie die bij de ene evaluatiestudie als 'best practice' uit de bus kwam, er in een volgende studie slecht vanaf komt. Het 'kookboekidee', dat in discussies over 'best practices' vaak opduikt, dient daarom sterk genuanceerd te worden.

Deze nuanceringsvinden we terug in de discussie die vanaf het begin jaren negentig onder aanvoering van Pawson en Tilley (zie o.a. *the British Journal of Criminology*, 1994 en *Evaluation*, 2002) gevoerd wordt. Het kernpunt van hun kritiek op het quasi-experimentele design (en de meta-evaluaties van deze studies zoals gepropageerd door de Campbell Collaboration) is dat de aan de interventie ten grondslag liggende theorie niet of nauwelijks in de analyse en discussie wordt betrokken. Volgens Pawson en Tilley zijn de uitkomsten van empirisch onderzoek alleen waardevol als we begrijpen welke mechanismen eraan ten grondslag liggen en welke contexten voorwaardelijk zijn voor het bestaan van een relatie tussen een interventie en een uitkomst. Om vooruitgang te boeken moeten de sociale interventies die geëvalueerd worden, gebaseerd zijn op een systematische integratie van achterliggende mechanismen en voorwaardelijke condities. Ook overzichtsstudies van effectevaluaties winnen aan belang wanneer niet alleen aandacht wordt besteed aan de 'klassieke' methodologische eisen, maar ook aan de relatie tussen context, mechanismen en uitkomsten van sociale interventies (zie Leeuw, Van Der Knaap & Bogaerts, *American Journal of Evaluation* (in press)).

Hier lijkt een rol weggelegd voor onderzoekers die zich tot nu toe vooral bezig hebben gehouden met 'correlationele' studies. Zij hebben vaak een sterk theoretische achtergrond en beschikken over kennis over de contexten waarin interventies plaatsvinden. Daarbij kan bijvoorbeeld gedacht worden aan demografische en leefstijlkenmerken en aan attitudes van specifieke bevolkingsgroepen. De uitkomsten van effectstudies kunnen overigens ook weer een belangrijke bijdrage leveren aan de voortgang van de theorievorming. Zij kunnen immers een zeer krachtige toetsing vormen van de algemenere theorieën, tenminste wanneer de opzet van de interventies en de metingen in de effectstudies grondig doordachte afleidingen uit de theorieën vertegenwoordigen.

Voor de hierboven geschetste synthese is misschien wel een omslag in het denken nodig en moet koudwatervrees worden overwonnen voor het gebruik van experimentele designs in de maatschappijwetenschappen. Daarbij is het goed ons te realiseren dat de maatschappij al vol zit met experimenten (interventies), maar dat wij er als onderzoekers vaak niet bij zijn om de effecten van die experimenten op de juiste manier te meten. Als reactie op het veel genoemde bezwaar dat sociale experimenten ethische problemen met zich brengen, kunnen we ons afvragen of het niet veel kwalijker is om als samenleving experimenten uit te voeren en er vervolgens niets van te leren. De kunst is om er als wetenschapper op tijd bij te zijn wanneer sociale interventies plaatsvinden. Zo kan een interventie zodanig worden opgezet dat deze een sterk theoretisch fundament heeft en dat een kwalitatief goede effectmeting mogelijk is.

De verantwoordelijkheid voor het scheppen van de juiste voorwaarden voor kwalitatief

hoogstaande en realistische effectevaluaties ligt bij beleidsmakers én wetenschappers. Zowel bij beleidsmakers in de publieke sector als bij onderzoekers begint dit besef meer en meer door te dringen. De hoop is dat de verbeterende contacten tussen de beleidswereld en wetenschappelijke onderzoekers de mogelijkheden voor goede effectevaluaties vergroten. Als wetenschappers in een vroeg stadium bij de planvorming betrokken raken, is de opzet van interventies op basis van (deductie uit) beschikbare kennis en theorieën meer gegarandeerd en zal de terugkoppeling van de uitkomsten van de effectevaluaties naar de theorie krachtiger en waarschijnlijker zijn (inductie). Er is op dit gebied echter nog veel terrein te winnen en in een tijd dat de roep om kennis over 'wat werkt en wat niet' zo groot is als nu, is het van groot belang dat deze kennis een stevige fundering houdt in kwalitatief hoogstaand wetenschappelijk onderzoek.

Tamar Fischer