

Veldexperimenten voor Beleid

Sander Gerritsen en Patricia Prüfer

In opdracht van het ministerie van OCW voert het Centraal Planbureau (CPB) regelmatig veldexperimenten uit. De auteurs geven een aantal voorbeelden van veldexperimenten die het CPB recentelijk heeft uitgevoerd en op dit moment uitvoert. Ze sluiten het stuk af met een paar aanbevelingen: (1) er moet een geloofwaardige controlegroep worden gevormd, (2) commitment van alle betrokken partijen bij het veldexperiment is noodzakelijk, (3) er moeten gegevens worden verzameld om het experiment te monitoren, (4) het is nuttig informatie te verzamelen over onderliggende mechanismen van het effect van het experiment, (5) het is nuttig om voorafgaande aan een gerandomiseerd experiment een poweranalyse uit te voeren.

1 Inleiding

Dit themanummer gaat over veldexperimenten. Dat uit dergelijke experimenten veel geleerd kan worden voor beleid, behoeft geen betoog. Dat is immers veelvuldig in de literatuur besproken (zie bijvoorbeeld Van Elk 2014; of Kooreman en Potters 2011). Ook willen we niet uitvoerig ingaan op welke econometrische technieken en methoden de onderzoeker tot zijn beschikking heeft om tot een goede evaluatie van een veldexperiment te komen. Daar zijn immers ook goede overzichtsartikelen voor (zie ook Van Elk 2014; Kooreman en Potters 2011). In dit stuk willen we vooral ingaan op de praktische kant van het doen van veldexperimenten. Waar loop je als onderzoeker soms tegenaan? Het CPB voert regelmatig veldexperimenten uit in opdracht van het ministerie van OCW. We zullen voorbeelden geven van veldexperimenten die het CPB recent heeft gedaan en van experimenten waar we op dit moment mee bezig zijn. Aan de hand van deze ervaringen sluiten we het stuk af met een paar aanbevelingen.

Deze ervaringen beslaan een vijftal experimenten, die worden uiteengezet in de komende paragrafen. Hierin worden zowel de geslaagde experimenten als een minder geslaagd experiment beschreven. In de volgende paragraaf beginnen we met het veldexperiment bij ROC Rijnijssel. Dit experiment betrof het ideaal voor de onderzoeker: het gerandomiseerde veldexperiment. In zo'n experiment worden mensen via loting aan een behandel- of controlegroep toegewezen. De behandelgroep krijgt de interventie, terwijl de controlegroep de interventie onthouden wordt. De loting zorgt ervoor dat voorafgaande aan het experiment de behandel- en controlegroep vergelijkbaar zijn, waardoor na afloop het effect van de interventie

kan worden bepaald door de uitkomsten van de behandelgroep met die van de controlegroep te vergelijken.¹

In paragraaf 3 leggen we vervolgens - aan de hand van een ander voorbeeld - uit dat een dergelijke opzet niet altijd noodzakelijk is om een interventie te evalueren, en dat er ook andere manieren zijn. In paragraaf 4 beschrijven we dan een experiment dat minder succesvol verlopen is om daarna, in paragrafen 5 en 6, experimenten te beschrijven waarbij de lessen van het minder geslaagde experiment zijn meegenomen. In paragraaf 7 volgt ten slotte de conclusie.

2 ROC Rijnijssel: intensieve begeleiding voor mbo 2-leerlingen

Een voorbeeld van een succesvol verlopen veldexperiment, is het experiment met intensieve begeleiding voor leerlingen op mbo-2 niveau op het ROC Rijnijssel in Arnhem (zie Van Elk et al. 2012). Dit experiment was gericht op het tegengaan van schooluitval. Aanleiding voor het experiment was de relatief hoge uitval die dit opleidingsniveau traditioneel kent, en de behoefte bij het ministerie van OCW om meer ‘evidence-based’ kennis te vergaren over de werking van interventies gericht op het reduceren van schooluitval. Het ministerie stelde voor dit experiment middelen beschikbaar voor een coach die intensieve begeleiding gaf aan enkele groepen van gemiddeld 20 mbo-studenten. Deze coach had als doel om uitval te voorkomen en begeleidde de studenten door middel van een-op-een gesprekken om de studievoortgang en persoonlijke problemen te bespreken. Daarnaast ondersteunden de coaches hen tijdens de reguliere lessen, organiseerden zij groepsactiviteiten en gingen zo nodig op huis- en stagebezoek. Enkele klassen binnen het ROC werden door middel van loting toegewezen aan een experimentgroep (met een extra coach) of een controlegroep (zonder extra coach). Op die manier kon het effect van de interventie worden bepaald door de uitval uit de opleiding in de experimentgroepen na 1 en 2 jaar te vergelijken met de uitval in de controlegroepen.

Uit het experiment bleek dat de leerlingen in de experimentgroep minder uitvielen dan de leerlingen in de controlegroep. Met andere woorden: de inzet van de coaches verkleinde de kans op schooluitval (Van Elk et al. 2012). Dit experiment was een mooi voorbeeld van een situatie waarbij beleidsmakers van het ministerie, onderwijsprofessionals op het ROC en onderzoekers intensief samenwerkten om te leren over de effecten van nieuw beleid. Dit project was daarmee misschien ook wel een schoolvoorbeeld van een experiment waarbij in grote lijnen alles volgens plan is uitgevoerd. De loting heeft naar behoren plaatsgevonden, dat wil zeggen dat na vaststelling van de loting alle ingelote opleidingen ook daadwerkelijk behandelgroep werden en alle uitgelote opleidingen controlegroep. Ook was er commitment

¹ Gestratificeerd loten verdient uiteraard de voorkeur, als dat mogelijk is. In dat geval kunnen op basis van beschikbare informatie voorafgaande aan het experiment ‘paartjes’ gemaakt worden van personen die op elkaar lijken wat betreft opleidingsniveau, geslacht etc. Vervolgens kan binnen die paartjes geloot worden.

van alle betrokken partijen, waardoor zowel de loting als de dataverzameling zonder problemen verliep. Verder was de opzet van het experiment zodanig dat de behandelgroep de controlegroep niet of nauwelijks kon beïnvloeden (de coach kon immers alleen zijn toegewezen groep coachen, en de leerlingen zaten in verschillende behandel- en controlegroepen).² Kortom, dit experiment was qua uitvoering redelijk goed verlopen. Maar betekent dit nu dat deze ideale opzet altijd mogelijk is? En moet dat altijd?

3 De Wijksschool

Nieuwe beleidsevaluaties hoeven niet alleen maar door middel van randomisatie plaats te vinden. Bij de invoering van nieuw beleid zijn er vaak ook andere kansen die voor evaluaties kunnen worden benut. Zo vinden er in het onderwijs vaak proefprojecten plaats, die goede mogelijkheden bieden voor een evaluatie. Dit kan bijvoorbeeld door de regels te gebruiken voor toewijzing aan de pilot (de experimentgroep) of de controlegroep. Soms kunnen deze regels ervoor zorgen dat een geloofwaardige controlegroep wordt gecreëerd. Dit laatste idee is gebruikt bij een evaluatie van de wijksscholen in Rotterdam (zie Van Elk et al. 2013). Dit type experimenten worden ook wel quasi-experimenten genoemd.³

In 2009 is een tweejarige pilot gestart met twee wijksscholen die jongeren met complexe problemen op meerdere leefgebieden een intensief programma biedt. De pilot had als doel hen te begeleiden naar het reguliere onderwijs of naar werk. Jongeren die aan een bepaald profiel voldeden werden door de gemeente aangemeld bij de Wijksschool. Door de beperkte capaciteit op de wijksscholen was na verloop van tijd echter geen plek meer voor nieuwe deelnemers. Dit bood de mogelijkheid een controlegroep te vormen. Deze kon namelijk worden geconstrueerd uit de groep jongeren die wel in aanmerking kwamen voor de Wijksschool, maar niet konden deelnemen aan het programma omdat er geen plek meer was op het moment dat zij zich aanmeldden. Als er wel plek was geweest, waren deze jongeren verwezen naar de Wijksschool. Deze groep van vergelijkbare jongeren kreeg vervolgens door de gemeente een regulier traject aangeboden.

Het CPB heeft deze toewijzingsprocedure gebruikt voor de evaluatie van de effecten van de Wijksschool. De effectiviteit is bepaald door sociaal-economische uitkomsten van jongeren op de Wijksschool te vergelijken met jongeren in een regulier

² Bij elk veldexperiment moet goed worden nagedacht over mogelijke ‘contaminatie’ of ‘besmetting’. Er is sprake van contaminatie als deelnemers in de controlegroep ook een deel van de interventie toebedeeld krijgen. Dat moet zoveel mogelijk voorkomen worden.

³ Soms kan een quasi-experiment zelfs te prefereren zijn boven een gerandomiseerd experiment. Bij gerandomiseerde experimenten kunnen de deelnemers weten dat zij onderdeel zijn van een onderzoek waarvan een bepaalde uitkomst in hun belang kan zijn. In dat geval kunnen ze hun gedrag daarop aanpassen, en worden misschien niet meer de effecten gemeten die de onderzoeker zou willen meten.

traject.⁴ Uit de analyse kwam naar voren dat leerlingen van de Wijkschool geen grotere kans op onderwijsdeelname of werk hadden vergeleken met de controlegroep. Daarnaast bleek dat een verwijzing naar de Wijkschool zelfs tot een toename in criminaliteit leidde ten opzichte van een verwijzing naar een regulier traject. Een verklaring hiervoor werd gevonden in het feit dat de Wijkschool, in tegenstelling tot de reguliere trajecten, jongeren bij elkaar zet die een crimineel verleden hebben. Hierdoor konden negatieve spillovers (het van elkaar leren van crimineel gedrag) ontstaan die leidden tot meer crimineel gedrag.

Dit was een verrassende uitkomst. De verwachtingen waren immers dat er positieve effecten van de Wijkschool zouden zijn; met die intentie was zij immers ook bedacht. Daarmee onderstreept dit onderzoek ook wel het belang van veldexperimenten. Als de Wijkschool niet eerst getest was in deze vorm, en het wijkschoolconcept nationaal geïmplementeerd was, dan zou dat waarschijnlijk tot maatschappelijke schade hebben geleid. Het loont dus om eerst met een idee of beleid te experimenteren, alvorens het landelijk uit te rollen.

Daarnaast laat dit veldexperiment zien dat als randomisatie niet mogelijk is, er slim gebruik kan worden gemaakt van de toewijzingsregels om een geloofwaardige controlegroep te vormen. Vaak dienen dit soort mogelijkheden zich aan wanneer er duidelijke restricties worden opgelegd aan deelname aan een pilot. In dit voorbeeld waren dat capaciteitsrestricties, maar men kan ook denken aan budgetrestricties (slechts een bepaalde hoeveelheid geld is beschikbaar, en wie het eerst komt wie het eerst maalt), of een kwaliteit- of niveaurestrictie, bijvoorbeeld als op basis van test scores of IQ studenten worden ingedeeld in een behandel- of controlegroep. Het veldexperiment met de Wijkschool kan dus, hoewel niet gerandomiseerd, gezien worden als een redelijk succesvol verlopen experiment. Helaas is dat niet altijd het geval, zoals bij het mbo-taalexperiment.

4 Mbo Taalexperiment

In dit experiment wordt gekeken wat de effecten zijn van het geven van extra taallessen aan mbo-leerlingen. Dit experiment is in het leven geroepen vanuit het ministerie van OCW omdat er zorgen waren – en nog steeds zijn – over het taalniveau van mbo-leerlingen. Taalvaardigheid is immers een belangrijke factor voor later arbeidsmarktsucces. Uit een aantal onderzoeken blijkt dat het taalniveau van een deel van de mbo-studenten achterblijft bij de niveaus die ze geacht worden te bereiken aan het eind van hun studie. Ook blijkt dat het niveau van geletterdheid onder de beroepsbevolking sinds het midden van de jaren negentig het meest is afgenomen onder middelbaar opgeleiden. De zorgen over taalniveaus van met name middelbaar opgeleiden hebben geleid tot meer beleidsaandacht voor taal, resulte-

⁴ Daarbij uiteraard rekening houdend met de verschillen in aanmeldingsdatum. In econometrisch jargon: er is een ‘regression discontinuity’ model gebruikt voor het schatten van de effecten.

rend in onder andere de invoering van centrale examens in het mbo voor lezen en schrijven vanaf 2014-2015.

Het taalexperiment is in drie achtereenvolgende schooljaren uitgevoerd, onder telkens een nieuw cohort eerstejaars mbo-leerlingen van niveau 3 en niveau 4. In elk schooljaar werd ongeveer de helft van de opleidingen ingeloot. Binnen deze opleidingen (de behandelgroep) werden extra lessen gegeven. De andere helft kreeg deze taallessen niet en vormde dus de controlegroep. Het aantal opleidingen dat aan het experiment zou meedoen werd voorafgaande aan het experiment middels een 'poweranalyse' bepaald. In een dergelijke analyse wordt een indicatie verkregen van de grootte van de steekproef, dat wil zeggen, het aantal waarnemingen dat nodig is om relevante effecten te kunnen vinden.⁵ Op basis hiervan zijn er in het eerste schooljaar 36 opleidingen (van de 72) ingeloot die gedurende het eerste studiejaar extra taallessen kregen. De overige 36 kregen deze extra taallessen dus niet.

De focus van de extra lessen lag op het aanleren van lees- en schrijfstrategieën. Voor de lessen was apart lesmateriaal ontwikkeld en zij werden gegeven door getrainde docenten. Beoogd was een uitbreiding van het aantal lessen met 30 uur, wat gemiddeld overeen zou komen met een uitbreiding van de onderwijstijd voor Nederlands met 35 procent. Aan het begin en het eind van het schooljaar werden toetsen afgenomen op het vlak van taalverzorging, lezen en schrijven.

Uit een (tussen)evaluatie⁶ van het CPB bleek dat de extra taallessen geen effect hebben gehad op de taalniveaus aan het eind van het schooljaar. Ook waren er geen aanwijzingen voor positieve effecten voor bepaalde subgroepen.⁷ De vervolgvraag die gesteld werd was waarom er geen effecten waren van de extra lessen. Een voor de hand liggende gedachte was dat het de extra lessen zelf waren die geen effect hadden gesorteerd. Maar uit nadere analyse, aan de hand van enquêtes en gesprekken met betrokkenen, kwam een ander beeld naar boven.

Diverse factoren leken een rol te hebben gespeeld voor het gevonden nul-effect, hoewel niet precies viel te ontleden in welke mate. Ten eerste de gerealiseerde omvang van de interventie. Het aantal daadwerkelijk gerealiseerde extra lessen was met gemiddeld 24 lessen 20 procent lager dan het vooraf beoogde aantal van 30 lessen. Dit betekende dus dat de interventie niet volledig was uitgevoerd. Een andere belangrijke factor leek de lage motivatie van studenten voor de extra lessen. Dit uitte zich onder meer in een hoog verzuim dat bijna twee keer zo hoog was als bij de reguliere lessen Nederlands. Andere factoren waren een ontbrekend belang van de toetsen voor de studievoortgang, ontevredenheid bij de taaldocenten met het ontwikkelde lesmateriaal en een gebrekkig draagvlak bij de opleidingen voor het experiment. Dit laatste leek vooral te maken te hebben met

⁵ Relevante effecten wil hier zeggen de minimale effectgrootte van de interventie welke de onderzoeker met zijn onderzoek wil aantonen. Hoe hoger de *power* van het experiment, hoe hoger de kans dat men kleine effecten van de interventie kan vinden.

⁶ Deze tussenevaluatie is niet openbaar gemaakt. In 2015 wordt het rapport opgeleverd met de resultaten van de afgelopen drie jaar.

⁷ Er is uitgesplitst naar geslacht en niveau.

het feit dat docenten van buiten de opleidingen de extra lessen verzorgden. Ten slotte kan de kwaliteit van de ingezette taaldocenten een rol hebben gespeeld: ongeveer 40 procent van hen was niet bevoegd om lessen Nederlands te verzorgen.⁸

Uit het experiment van dit eerste schooljaar bleek dus dat de implementatie van de extra taallessen niet optimaal is verlopen. Extra taallessen zouden misschien wel hebben kunnen helpen, maar niet in deze vorm. Wat nu als de docenten wel tevreden waren over het lesmateriaal en er een groot draagvlak was voor het geven van de extra taallessen? Waren er dan wel effecten van de extra taallessen gevonden?

Dit project heeft laten zien hoe belangrijk monitoring van een veldexperiment is. Dit wil zeggen dat wordt onderzocht in hoeverre het beleid of de interventie waarvan het effect moet worden gemeten, ook daadwerkelijk volgens plan is uitgevoerd. Voor een goede effectevaluatie van de taallessen zou op zijn minst aan een paar randvoorwaarden moeten zijn voldaan, zoals draagvlak onder de docenten voor het geven van de extra lessen en tevredenheid over het lesmateriaal.

Daarnaast laat het ook zien dat het belangrijk is informatie te verzamelen over de context waarbinnen het wordt uitgevoerd. Het hoge verzuim van de studenten en de lage motivatie voor de taallessen gaven een extra indicatie waarom de extra taallessen niet tot effecten hebben geleid.

Inmiddels heeft het CPB het derde experimentjaar afgerond, waarbij de lessen die zijn getrokken uit het eerste jaar zijn meegenomen. De taalinterventie is in dit derde jaar beter van de grond gekomen. Docenten zijn tevreden met het nieuw ontwikkelde lesmateriaal, en er is veel minder verzuim onder studenten.

Helaas heeft zich in dit experimentjaar echter een ander probleem voorgedaan, dat vooral lag in de methodische opzet van het experiment. In het derde experimentjaar is namelijk de loting niet naar behoren uitgevoerd. Van de 83 opleidingen waarover zou worden geloot (42 in de behandel- en 41 in de controlegroep) zijn ongeveer 30 opleidingen uitgevallen. Dit had onder andere te maken met het feit dat de opleidingen uit de onderzoeksgroep geen toestemming van de teamleider hadden gekregen om deel te nemen aan het experiment (dit terwijl het hogere management die goedkeuring wel had gegeven). Ook waren er docenten die op eigen houtje hadden besloten voor sommige ingelote groepen de interventie niet uit te voeren. Daarnaast zaten er van de overgebleven opleidingen een paar opleidingen bij die zich niet helemaal aan de loting hielden. Sommige opleidingen waren ingeloot in de behandelgroep, maar werden controlegroep en kregen dus geen extra lessen. Sommige opleidingen die waren uitgeloot kregen deze extra lessen alsnog en werden dus behandelgroep. In de literatuur wordt dit gemarchandeerd met controle- en behandelgroep ook wel met de term ‘non-compliance’ aangeduid. Dat hoeft in principe geen probleem te zijn omdat het kan worden opgelost door een statistische techniek die bekend staat als instrumentele

⁸ Onbevoegd wil hier zeggen dat de docenten geen eerste- of tweedegraadsbevoegdheid Nederlands hadden. Ze waren door het betreffende ROC wel als bekwaam bestempeld.

variabele analyse.⁹ Maar in dit geval was zelfs dat niet meer mogelijk aangezien de non-compliance daarvoor te groot was (dat wil zeggen dat er te veel opleidingen waren die zich niet meer aan de loting hadden gehouden).¹⁰

Een belangrijke les uit dit taalexperiment is dan ook dat commitment van alle betrokken partijen belangrijk is voor een goede uitvoering van het experiment. Alleen een goedkeuring voor het experiment, afgegeven door het hogere management, kan onvoldoende zijn, aangezien dit geen garantie is dat het lagere management en/of de werknemers op de werkvloer daar gehoor aan geven. In dit voorbeeld gingen sommige teamleiders hun eigen gang. Commitment van alle partijen, en misschien nog wel het meest van de degenen die direct of het dichtst bij het experiment betrokken zijn, is dus zeer belangrijk. De ervaringen met dit veldexperiment zijn meegenomen bij de opstart van een viertal gerandomiseerde veldexperimenten die het CPB uitvoert om voortijdig schoolverlaten (VSV) in het mbo tegen te gaan. Zo is bijvoorbeeld, voorafgaande aan de experimenten, de voorwaarde gesteld dat de subsidie voor het experiment alleen wordt verstrekt als de betrokken organisatie akkoord is met loting van deelnemers in behandel- en controlegroepen. Daarnaast is er bij elk ROC een (contact)persoon aangesteld die verantwoordelijk is gemaakt voor het experiment. Deze persoon zorgt ervoor dat de loting naar behoren wordt uitgevoerd en helpt bij de dataverzameling. We zullen twee van deze experimenten kort toelichten.

5 VSV-experiment: de verbeterde intake

Het CPB evalueert een veldexperiment bij een ROC waarbij de interventie bestaat uit een verbeterd intakegesprek. Dit gesprek heeft als doel de uitval van – vooral – risicoleerlingen te verminderen. In het veldexperiment wordt de verbeterde intake afgezet tegen de bestaande intake. Een intakegesprek is een gesprek dat een mbo-leerling aan het begin van de opleiding volgt, en dat door de studieloopbaanbegeleider wordt uitgevoerd. Zo'n gesprek moet ervoor zorgen dat risicoleerlingen worden herkend en dat leerlingen op een traject/opleiding terecht komen dat goed bij hen past. Dat zou uitval moeten voorkomen.

De oude intake bij het betreffende ROC was erg versnipperd over de opleidingen. Zij was niet gestandaardiseerd, varieerde in duur, en ook in type bijeenkomst (soms een groepsgesprek, soms individueel). Daarnaast werd er relatief weinig achtergrondinformatie over de leerling opgevraagd om een gedegen advies te kunnen geven. Zo werd er bijvoorbeeld geen informatie ingewonnen bij de ouders. Het idee was daarom opgevat om de intake te verbeteren en te standaardiseren.

⁹ In het regressiemodel zou hierbij de dummy die aangeeft of een opleiding extra taallessen heeft gekregen geïnstrumenteerd worden met de oorspronkelijke loting.

¹⁰ In econometrische termen: de 'first stage' was te zwak, dat wil zeggen dat de dummy die aangeeft of in de opleiding extra lessen zijn gegeven matig correleert met de oorspronkelijke lotingsdummy.

De nieuwe intake bestaat uit een individueel gesprek dat tussen de 45 en 60 minuten duurt en dat volgens een vast format verloopt. Daarbij wordt veel informatie opgevraagd over de student, en wordt ook informatie ingewonnen bij de ouders. Ook zijn de studieloopbaanbegeleiders voor de verbeterde intake geschoold in het herkennen van risicoleerlingen door middel van een driedaagse training. Het doel van de nieuwe intake is dat risicoleerlingen eerder worden herkend waardoor voortijdig schoolverlaten onder mbo-leerlingen kan worden teruggedrongen. In het veldexperiment zijn ongeveer 750 studenten geloot over de behandelgroep en de controlegroep. In elke groep zitten ongeveer 375 studenten. De leerlingen in de behandelgroep ondergaan de verbeterde intake, terwijl de leerlingen in de controlegroep nog de bestaande intake ondergaan. In 2016 zal het CPB met de uitkomsten van dit onderzoek komen. Daarbij zal ook worden ingegaan op mogelijke mechanismen die tot wel of geen effect (op voortijdig schoolverlaten) hebben geleid. Dit wordt o.a. gedaan door gebruik te maken van aanvullende informatie uit enquêtes die zijn uitgezet onder de studieloopbaanbegeleiders.

6 VSV-experiment: steviger verzuimbeleid

Op een ander ROC evalueert het CPB een veldexperiment waarbij wordt gekeken naar de effecten van nieuw verzuimbeleid op voortijdig schoolverlaten en verzuim. In het experiment wordt nieuw verzuimbeleid (behandelgroep) afgezet tegen het bestaande verzuimbeleid (controlegroep).

Het bestaande beleid is op dit moment zo vormgegeven dat deelnemers tussen de 18 en 23 jaar bij minstens 16 uur ongeoorloofd verzuim in 4 aaneengesloten weken worden gemeld bij een verzuimloket. Dit loket registreert de deelnemer, waardoor er contact wordt opgenomen met het Regionale Meld- en Coördinatiepunt voor voortijdig schoolverlaters (RMC). Dit RMC heeft de wettelijke verplichting om jongeren zonder startkwalificatie te volgen totdat ze 23 jaar zijn. Een maatschappelijk werker vanuit het RMC begeleidt de deelnemer vervolgens naar school, werk of hulpverlening.¹¹

Het nieuwe verzuimbeleid is anders vormgegeven. Het is erop gericht om eerder en steviger in te grijpen bij ongeoorloofd verzuim. Voordat de deelnemer wordt gemeld bij het verzuimloket (bij minstens 16 uur ongeoorloofd verzuim in 4 aaneengesloten weken), wordt met de deelnemer al het gesprek aangegaan over zijn/haar verzuim bij minstens 8 uur ongeoorloofd verzuim in 4 aaneengesloten weken. De deelnemer wordt opgeroepen voor een preventief spreekuur en krijgt zelfs een huisbezoek door RMC-medewerkers als hij/zij na 2 oproepen voor het spreekuur nog niet verschenen is. Daarnaast worden de ouders op de hoogte gesteld van het verzuim.

¹¹ Daarnaast blijft de mentor vanuit het ROC in contact met de deelnemer en biedt begeleiding, zodat voortzetten van de opleiding, dan wel terugkeer naar school mogelijk is

Het verschil tussen de behandel- en controlegroep is dus dat er in de behandelgroep eerder wordt ingegrepen (al bij 8 uur), dat de ouders daarvan op de hoogte worden gesteld, dat er een preventief spreekuur is, en dat een huisbezoek volgt als een deelnemer 2 keer niet op spreekuur is verschenen.

Voor dit veldexperiment heeft het CPB ongeveer 6000 leerlingen geloot over twee regimes: 3000 leerlingen in een behandel- en 3000 in een controleregime. Als een leerling minstens 8 uur verzuimt in het behandelregime krijgt zij te maken met het nieuwe verzuimbeleid. Doet een leerling dit in het controleregime, dan gebeurt er nog niks. Pas als de leerling 16 uur of meer verzuimt, wordt er ingegrepen en krijgt zij met het oude verzuimbeleid te maken. Het experiment is dus zo opgezet dat het totale effect van het regime bepaald kan worden door een vergelijking te maken tussen de leerlingen in de behandel- en controlegroep.

Dit is echter niet het enige wat we willen weten. In beide regimes zitten ook veel leerlingen die nog geen 8 uur verzuimen. We willen daarom ook weten wat de effecten zijn van het nieuwe verzuimbeleid voor degenen die ook daadwerkelijk 8 uur of meer verzuimd hebben (en dus naar spreekuur moesten). Vallen zij minder snel uit en verzuimen zij minder dan een vergelijkbare groep die nog onder het oude beleid valt? Om dit te weten te komen worden zij vergeleken met een controlegroep van leerlingen die minstens 8 uur verzuimd hebben maar met het oude verzuimbeleid te maken hebben gekregen.¹² Een vergelijking tussen deze groepen levert dan het gewenste effect op. Op die manier schatten we dus twee effecten: het totale effect van het ingeloot zijn in het nieuwe regime, en de effecten van het nieuwe verzuimbeleid voor diegenen die minstens 8 uur verzuimen (en dus naar spreekuur moesten). In 2016 zal het CPB met de uitkomsten van dit onderzoek komen. Daarbij zal ook worden ingegaan op mogelijke mechanismen die tot wel of geen effect hebben geleid. Dit wordt onder andere gedaan door gebruik te maken van aanvullende informatie uit enquêtes en gesprekken met degenen die het preventieve spreekuur hebben afgenomen in het nieuwe verzuimbeleid. De gesprekken dienen als ‘*anecdotal evidence*’ voor mogelijke mechanismen waarlangs het nieuwe verzuimbeleid effect heeft gesorteerd.

7 Conclusies

Welke lessen kunnen wij nu trekken uit deze experimenten? Volgens ons ten minste vijf.

Ten eerste is het belangrijk dat bij een veldexperiment een geloofwaardige controlegroep wordt gevormd. Het ideaal is uiteraard door middel van (gestratificeer-

¹² De behandelgroep zou hier kunnen verschillen van de controlegroep als de studenten, wetende dat ze in het behandelregime zitten, direct anticiperen op het nieuwe verzuimbeleid, bijvoorbeeld door minder te gaan verzuimen. In dat geval kan het percentage leerlingen dat minstens 8 uur verzuimt verschillen tussen behandel- en controlegroep. Echter, dit lijkt hier niet een heel groot probleem te zijn, aangezien de leerlingen niet weten in welk regime ze zitten. Dat is niet van tevoren aangekondigd.

de) randomisatie, zoals het geval was bij het experiment in Rijnijssel Arnhem, maar dit hoeft niet per se. Er zijn ook andere manieren om te evalueren. Bij de invoering van nieuw beleid zijn er vaak kansen die voor evaluaties kunnen worden benut. Zo vinden er in het onderwijs bijvoorbeeld vaak proefprojecten plaats, die goede mogelijkheden bieden voor een evaluatie. Dit kan bijvoorbeeld door de regels te gebruiken voor toewijzing aan de pilot (de experimentgroep) of de controlegroep. Soms kunnen deze regels er namelijk voor zorgen dat een geloofwaardige controlegroep wordt gecreëerd, zoals in het voorbeeld van de evaluatie van de wijkscholen in Rotterdam. Een aandachtspunt daarbij is dat ‘contaminatie’ of ‘besmetting’ zoveel mogelijk wordt voorkomen, dat wil zeggen dat alleen de behandelgroep de interventie krijgt toebedeeld (en niet ook de controlegroep). Dit is belangrijk, omdat dat anders het effect van de interventie kan vertekenen. Bij het experiment bij Rijnijssel was de opzet zo vormgegeven dat de kans daarop erg klein was. Dit gold overigens ook voor het veldexperiment met de wijkscholen, omdat het onwaarschijnlijk was dat de deelnemers die naar de reguliere trajecten werden gestuurd de deelnemers in de Wijkschool konden beïnvloeden.

Ten tweede is commitment van alle betrokken partijen bij het veldexperiment belangrijk, vooral van degenen die direct bij de uitvoering van het experiment betrokken zijn. Vaak zijn dat lagere managementlagen of werknemers op de werkvloer. Het contact en de afspraken die met hen kunnen worden gemaakt zijn misschien nog veel belangrijker dan die met een directeur. Als het hogere management goedkeuring aan een project heeft gegeven, is dat nog geen garantie dat het experiment goed zal verlopen. Een aanbeveling zou kunnen zijn om, als randvoorwaarde voor de subsidieverstrekking voor het experiment, het welslagen van de loting contractueel te laten vast leggen. Daarnaast zou het helpen om, binnen de organisatie waar het veldexperiment zich afspeelt, één persoon verantwoordelijk te maken voor het verloop van het experiment. Hij of zij moet dan wel bij machte zijn om de mensen die betrokken zijn bij het experiment aan te sturen. Dit is onder ander gedaan bij de twee genoemde veldexperimenten om voortijdig schoolverlaten tegen te gaan.

Ten derde is het nuttig bij een experiment gegevens te verzamelen voor de monitoring van het experiment. Dat wil zeggen dat over het verloop en de implementatie van de interventie informatie wordt verzameld waarmee kan worden onderzocht of de interventie ook daadwerkelijk volgens plan is uitgevoerd. In het voorbeeld van het taalexperiment is deze informatie verzameld en nuttig gebleken. Hieruit bleek namelijk dat extra taallessen niet helemaal volgens plan waren doorgevoerd. Er was onder andere ontevredenheid van de taaldocenten over het ontwikkelde lesmateriaal en een gebrekkig draagvlak voor het experiment.

Ten vierde, en aansluitend op het derde punt, is het nuttig informatie te verzamelen die iets zouden kunnen zeggen over onderliggende mechanismen van het effect. Een effectmeting is mooi, maar als niet duidelijk is waarom er wel of geen effect is, dan wordt wellicht weinig lering getrokken uit het experiment. Deze aanvullende informatie kan via enquêtes van betrokkenen bij het experiment worden verkregen en/of via zogeheten *anecdotal evidence*, waarbij middels gesprek-

ken een idee wordt gevormd over de onderliggende mechanismen. Dit wordt onder andere gedaan bij de veldexperimenten die het CPB uitvoert om voortijdig schoolverlaten tegen te gaan.

Ten vijfde is het nuttig voorafgaande aan een gerandomiseerd experiment poweranalyses uit te voeren. Zij geven een indicatie van de grootte van de steekproef die nodig is om relevante effecten te vinden. Daarbij is een aandachtspunt dat men ook rekening houdt met uitval van deelnemers. Het voorbeeld van het taalexperiment liet zien dat een behoorlijk deel kan uitvallen.

Deze vijf punten bieden waarschijnlijk geen uitputtende lijst waar onderzoekers rekening mee kunnen houden bij het opzetten van veldexperimenten. Elk onderzoek is immers weer anders. Maar zij kunnen de onderzoekers misschien wel op weg helpen.

Auteurs

Sander Gerritsen (e-mail: s.b.gerritsen@cpb.nl) is wetenschappelijk medewerker bij het Centraal Planbureau en Patricia Prüfer (e-mail: p.prufer@uvt.nl) is senior onderzoeker bij CentERdata, Tilburg University.

Literatuur

- Elk, R. van, 2014, Evidence-based beleid, in: R. Klaus, L. Borghans en I. Waterreus (eds), *Wat is Goed Onderwijs? Bijdragen uit de Onderwijseconomie*.
- Elk, R. van, M. van der Steeg en D. Webbink, 2012, Zorgt intensieve coaching voor minder voortijdig schoolverlaten? Empirisch bewijs van een gerandomiseerd experiment, CPB Discussion Paper 224.
- Elk, R. van, M. van der Steeg en D. Webbink, 2013, The effects of a special program for multi-problem school dropouts on educational enrolment, employment and criminal behaviour, CPB Discussion Paper 241.
- Kooreman, P. en J. Potters, 2011, De gouden standaard: Veldexperimenten in de voorbereiding en evaluatie van beleid, *TPEdigitaal*, vol. 5(3): 76-90.