

Op 15 april 2008 vond in Brussel de eerste ontbijtsessie van het Vlaams Evaluatieplatform plaats. De gastspreker was **Christel Scholten** van 'Research voor Beleid'. 'Research voor Beleid' is een onderzoeksinstelling die deel uitmaakt van de panteia-groep (met daarin ook EIM, NEA, IOO, Stratus en IPM). *Christel Scholten* is sociaal beleids-wetenschapper met een pak ervaring in beleidsevaluatief onderzoek in de domeinen zorg en welzijn, arbeidsmarkt, sociale zekerheid en onderwijs. De sessie werd bijgewoond door 22 deelnemers.

Onderwerp van de uiteenzetting was "**evalueren van beleid**". Er werd concreet ingegaan op de situatie in Nederland, maar vooral op 'Hoe beleid evalueren?'.

 **Research voor Beleid**
onderzoek van Brussel

Doel vanochtend: Hoe evalueer je beleid?

- Typen evaluatie
- (Quasi) experimenteel
- Realistic Evaluation (Pawson)
- Best of both worlds
- Tips bij evalueren
- BTW case



Wat de **situatie in Nederland** betreft, is er een groot aanbod aan instanties die beleidsevaluatief onderzoek doen. Ook de vraag bij beleidsmakers is groot. Beleidsevaluatie zit bij Nederlandse beleidsmakers meer en meer 'tussen de oren'. Aanwijzingen daarvoor zijn onder andere het VBTB-programma (van beleidsbegroting tot beleidsverantwoording), de activiteiten van de rekenkamers, en het feit dat wetgeving standaard om de vier jaar moet worden geëvalueerd. De traditie in Nederland is dat veel evaluatieonderzoek wordt uitbesteed omdat (ex post) evaluaties die gebeuren door de instantie die zelf wordt geëvalueerd of verantwoordelijk is voor het gevoerde beleid, niet serieus worden genomen. De bevoegde overheidsinstanties worden wel nauw betrokken bij evaluatieonderzoek, typisch in begeleidingscomités. Daarnaast zijn er uiteraard specifieke overheidsdiensten die vanuit een meer

onafhankelijke positie de taak hebben om te evalueren.

Wanneer we het hebben over **evalueren van beleid**, is het volgens Scholten belangrijk een onderscheid te maken tussen verschillende soorten. Vooral het onderscheid tussen ex ante en ex post evaluatie is van belang.

 **Research voor Beleid**
onderzoek van Brussel

Typen evaluatie

Voor
ex-ante evaluatie
impact assessment
explorative onderzoek
scenario studies
behoeftenpeiling

Tijdens
monitoring
proces evaluatie
tussentijdse evaluatie

Achteraf
ex-post evaluation (efficiency and effectiviteit)
kosten-baten analyse
best practices



Ex ante evaluatie vloeit voort uit de trend naar meer 'evidence based policy', d.w.z. eerst evidence, dan policy. Typisch bij ex ante evaluatie worden dan van minstens 2 beleidsalternatieven op systematische wijze de verwachte kosten en effecten vergeleken om tot gefundeerde keuze van een van de alternatieven te komen.

Ex post evaluatie heeft betrekking op uitgevoerd of in uitvoering zijnd beleid. De focus ligt in de meeste gevallen op de evaluatie van de beleidseffectiviteit. Ex post evaluatie is de meest bediscussieerde vorm, en ook de moeilijkste evaluatie gelet op multicausale verbanden en de dikwijls dynamische beleidsomgeving waarbinnen het beleid wordt uitgevoerd.

Ex post evaluatie is lange tijd – sinds de jaren '60 – vooral uitgevoerd onder de vorm van **(quasi) experimenteel onderzoek**. Daarbij wordt beoogd de mate van effect van een bepaalde maatregel in kaart te brengen. Kenmerkend voor die aanpak is het zgn. counterfactual (wat als de maatregel niet was genomen?), de sterk kwantitatieve invalshoek, de aselechte verdeling van de doelgroep van de maatregel over 2 groepen (een controlegroep en een experimentgroep), en de 'double blind' werkwijze (leden van de controlegroep en experimentgroep

weten zelf niet in welke groep ze zitten, en ook onderzoekers weten het niet). Beleidseffecten worden gemeten voor en na invoering van het beleid, als het verschil in uitkomst tussen de controlegroep en de experimentgroep waarbij men gedurende het experiment de overige condities gelijk houdt.

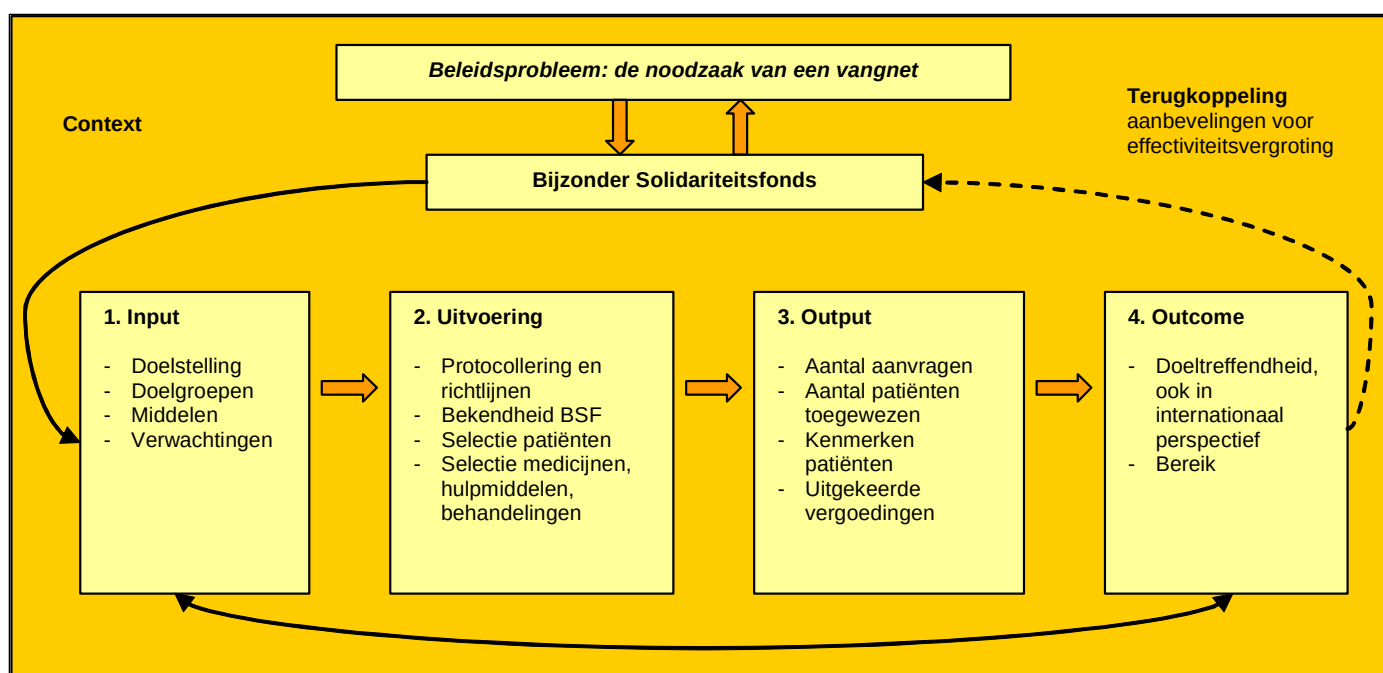
In de praktijk wordt zelden voldaan aan de zgn. gouden standaard (double blind) en moet het 'counterfactual' achteraf worden geconstrueerd (historische tijdreeks, grotendeels gelijkende groepen vergelijken). Scholten illustreerde dit met het voorbeeld van onderzoek naar de effectiviteit van de subsidies die scholen in Nederland ontvangen wanneer zij meer dan 70% allochtone leerlingen tellen. Het is in de praktijk moeilijk aan de vereisten van het experimentele design te voldoen (vandaar 'quasi'): dikwijls is het niet goed mogelijk effecten van beleid vast te stellen, ontbreekt een nulmeting of is nulmeting onmogelijk omdat nieuw beleid of wetgeving enkel de bevestiging is van een bestaande praktijk, zijn er problemen met de deelnemers aan de experimenten (ongemotiveerden, uitval) of is een controlegroep waar beleid geen betrekking op heeft moeilijk te vinden.

Pawson maar ook anderen hebben sinds 1997 kritiek geuit op het (quasi) experimenteel onderzoek. De kern van hun kritiek is dat het enkel uitkomsten geeft maar geen verklaringen. Men kijkt naar inputs en outputs, maar wat daartussen gebeurt blijft een 'black box'. De focus ligt op de interne validiteit (constructie en onderhoud van het experiment en de controlegroep) maar dit zegt niets over externe validiteit (verschillende onderzoeken over hetzelfde beleid komen tot verschillende conclusies). M.a.w.: als de context niet wordt betrokken, is geen generaliseerbaarheid mogelijk. Effecten op het terrein kunnen er zijn dankzij of ondanks het gevoerde beleid.

Het alternatief van Pawson ('**realistic evaluation**') start vanuit de 'beleidstheorie', d.w.z. de expliciete of impliciete aannames van de beleidsmakers omtrent de werking en effecten van de beleidsmaatregel die wordt onderzocht (relatie tussen de inzet van middelen en beoogde doelstellingen). Vanuit die beleidstheorie stelt men de vraag onder welke omstandigheid voor wie het beleid zou kunnen werken. Context, werkende mechanismen en outcome worden in hun samenhang blootgelegd.

Bij *realistic evaluation* worden dus expliciet vragen gesteld naar de probleemstelling of aanleiding voor de maatregel, de andere opties die werden bekeken bij de selectie van de maatregel, de overwegingen die hebben meegespeeld bij de uiteindelijke keuze, de uitvoering van de maatregel (Wordt de maatregel uitgevoerd? Werkt de maatregel als beoogd? Waar wel, waar niet? Waarom? Zijn er mechanismen die de maatregel blokkeren of anders doen werken dan vooraf bedacht?...) en het effect van de maatregel.

In onderzoek dat door *Research voor Beleid* wordt uitgevoerd, wordt de aanpak van Pawson twee maal doorlopen (zie onderstaand schema met een voorbeeld voor het bijzonder solidariteitsfonds). De eerste keer om de beleidstheorie te reconstrueren: wat was het probleem of de aanleiding? Wat is de relatie tussen de aanleiding en het ingezette middel? Enz. Doel van deze theoretische fase is het beleid toetsbaar te maken. De ervaring is dat er dikwijls zeer weinig informatie is over dit soort vragen, of dat blijkt dat er weinig over werd nageacht. Na reconstructie van de beleidstheorie wordt het schema een tweede keer doorlopen, teneinde via onderzoek en veldwerk na te gaan hoe de maatregel werkt in de praktijk.



De ervaring hier is dat relaties steeds moeilijker aan te tonen zijn naarmate men in het schema naar rechts opschuift (input, uitvoering, output, outcome). Er is dus veeleer sprake van plausibiliteit dan van causaliteit.

In de praktijk wordt dus veelal gewerkt met een **combinatie** van (quasi) experimenteel onderzoek en realistische evaluatie. Een goede evaluatie is zowel kwantitatief als kwalitatief, en geeft zowel effecten als verklaringen. Enkele vuistregels zijn: start met beleidsreconstructie (wat is het beleid - doelen, instrumenten, causaliteit?), baseer daar het evaluatiekader op, betrek verschillende bronnen, vergeet de counterfactual niet en zoek beleidsmakers die open staan voor leren (spiegel voorhouden).

Scholten ging tot slot in op het concrete **voorbeeld** van een EU experiment waarbij men de BTW wenste te verlagen voor arbeidsintensieve diensten om de werkgelegenheid voor laaggeschoolde arbeid te stimuleren. Scholten legde eerst uit wat hier de beleidstheorie was: door verlaging van de BTW daalt consumentenprijs, waardoor de vraag naar en daarmee de werkgelegenheid voor diensten zoals kappers stijgt. Vervolgens werd toegelicht hoe de beleidsevaluatie werd opgezet: het doortrekken van de bestaande trend in werkgelegenheid werd als counterfactual genomen, het nagaan van de causaliteit van de maatregel gebeurde aan de hand van een vergelijking van prijzen van kappers voor en na de invoering maatregel alsook van de vraag naar die diensten en gerelateerde werkgelegenheid. Als controlegroep werden vergelijkbare diensten genomen, zoals schoonheidsspecialisten, waarop de BTW-maatregel niet van toepassing was. Er werd ook trendanalyse van de werkgelegenheid uitgevoerd. Tot slot lichtte Scholten de resultaten van deze evaluatie toe: in de kappersbranche steeg werkgelegenheid kort, maar viel daarna terug op de voorspelde trend. Uit de enquête bij kappers bleek dat de BTW verlaging slechts beperkt aan klanten werd doorgegeven; de beleidstheorie ging dus voor een groot deel niet op. De conclusie van het onderzoek was dat het niet waarschijnlijk is dat de maatregel het beoogde effect heeft gehad, namelijk heeft geleid tot een verhoging van de werkgelegenheid bij de doelgroep van diensten.

Tijdens de informele gedachteswisseling na de uiteenzetting van Scholten werden heel wat thema's aangesneden:

- De opvolging van de resultaten van beleidsevaluatief onderzoek;
- De rol van doelgroepen en andere actoren bij het opzetten en uitvoeren van beleidsevaluaties, en mogelijke problemen van overbevraging en evaluatiemoetheid;
- De voor- en nadelen van uitbesteden van onderzoek door de administratie aan onderzoeksinstellingen, de sterktes en zwaktes van onderzoeksinstellingen ten opzichte van universiteiten, en de vraag naar capaciteit langs de kant van de overheid om onderzoek uit te voeren / goed uit te besteden;
- De kostprijs van evaluaties en de prioriteitenkeuze (kosten-baten verhouding van individuele beleids-evaluaties, selectie van te evalueren beleidsmaatregelen);
- De mate waarin onderzoekers niet enkel onderzoeksconclusies maar ook beleidsaanbevelingen kunnen/mogen formuleren;
- De doorlooptijd van evaluatieonderzoek als dat inderdaad grondig en in twee fasen gebeurt;
- De problemen die men in de praktijk tegenkomt bij de reconstructie van de beleidstheorie;
- De interactie en verhoudingen tussen evaluator en opdrachtgever gedurende het evaluatieproces, en ook de mogelijke beïnvloeding van het onderzoek en de conclusies door de opdrachtgever;
- De wenselijkheid van evaluatiestandaarden of protocols om de kwaliteit van beleidsevaluaties te waarborgen en een gelijk speelveld te krijgen tussen concurrenten;
- De verschillen tussen Nederland en Vlaanderen qua organisatie en werking van de overheid en de implicaties daarvan voor beleidsevaluatie;
- De (miskenning van) impliciete vormen van beleidsevaluatie die gebeuren door onder andere gemeenteraden en parlementen;
- De meerwaarde en beperkingen van sterk kwantitatief onderzoek t.o.v. meer kwalitatief onderzoek;
- Het relatieve belang dat men zou moeten hechten aan ex ante dan wel ex post evaluatie.

Naar aanleiding van de evaluatie van de ontbijtsessie werden door de deelnemers enkele mogelijke thema's voor een volgende ontbijtsessie opgesteld:

- Beleidsevaluatie in Vlaanderen: een overzicht en stand van zaken;
- Een recente Vlaamse beleidsevaluatie (een voor iedereen bekende, concrete situatie) – bespreking van een case – specifieke praktijkvoorbeelden;
- Een panel met beleidsmakers over wat ze met beleidsevaluaties doen, wat wordt door hen geapprecieerd en wat niet...;
- Doorwerking van beleidsevaluatieve studies in beleid: waarom wel/niet; hoe verbeteren?
- Technieken van impactanalyse;
- De situatie op het vlak van beleidsevaluatie in UK, F, D...;
- Actoren uit de Europese evaluatiepraktijk aan het woord laten;
- Evaluatieonderzoek bij de FOD Justitie en Binnenlandse zaken
- De kostprijs van beleidsevaluaties, de afweging van kosten en baten van beleidsevaluaties.