

# **EVALUEREN IS LEREN**

*Een studie naar de kwaliteit van beleidsevaluatieonderzoek*

Erik Spoelstra  
Rijksuniversiteit Groningen  
Faculteit der Ruimtelijke Wetenschappen  
Technische Planologie  
september 2004

## Samenvatting

Beleidsvaluatie is de laatste jaren in Nederland in opkomst. Doel hiervan is om het functioneren van een beleid te beoordelen. Hier zitten echter nogal wat haken en ogen aan. Dit onderzoek is gericht op enkele vragen die hiermee verband mee houden, namelijk:

*Waar ligt de kwetsbaarheid van beleidsvaluatieonderzoek en hoe kan er een methode worden ontwikkeld om beleidsvaluatieonderzoek op een efficiënte manier te kunnen beoordelen en te verbeteren en kan hiermee ook het opstellen van beleid worden verbeterd?*

Het hieruit voortvloeiende doel van dit onderzoek is drieledig:

1. Het in kaart brengen van de kwetsbaarheid van beleidsvaluatieonderzoek;
2. Het ontwikkelen van een methodiek voor het bepalen van de kwaliteit van beleidsvaluatieonderzoek;
3. Het toepasbaar maken van de methode voor het verbeteren van beleidsvaluatieonderzoek én beleid.

Om deze drie doelen te volbrengen is de allereerst een theoretisch kader gehanteerd, waarin de begrippen beleid, beleidsonderzoek en evaluatie nader zijn toegelicht. Beleid moet voldoen aan drie eisen: effectiviteit, efficiëntie en legitimiteit. Hiervoor zijn verschillende onderzoekstypen in omloop, die allemaal dienen voor uitvoeren van beleidsvaluatieonderzoek. Gebleken is dat de keuze voor een onderzoekstype niet makkelijk te maken is. Uit de literatuur blijkt dat een evaluatieonderzoek daarnaast niet altijd volgens de regels wordt uitgevoerd. Ook blijkt verwarring te bestaan bij de opdrachtgevers en onderzoekers over de typen onderzoek. Daarom is een driedeling aangebracht in evaluatieonderzoek: monitoring, auditing en evaluatie. Deze drie onderzoekstypen houden verband met de genoemde begrippen effectiviteit, efficiëntie en legitimiteit. Maar het onderzoek is hiermee nog niet afgerond.

Om de kwaliteit van een evaluatieonderzoek te bepalen dient een zogenaamde meta-evaluatie te worden uitgevoerd. Dit houdt simpel gezegd de evaluatie van een evaluatie in. De methodiek voor meta-evaluatie is als input gebruikt voor het opstellen van een methode van onderzoek. Stufflebeam (1974) heeft een checklist opgesteld met daarin de eisen waaraan een evaluatie zou moeten voldoen. De afgelopen decennia zijn ook in Nederland onderzoekers geweest die een methode hebben aangereikt voor het uitvoeren van een meta-evaluatie. Om te komen tot een methodiek voor dit onderzoek is de checklist van Stufflebeam gecombineerd met de 'Nederlandse' methoden. Met de in dit onderzoek opgestelde methodiek wordt allereerst een poging gedaan om de kwetsbaarheid van beleidsvaluatieonderzoek te brengen. Het ontwikkelen van een methodiek voor het bepalen van de kwaliteit van beleidsvaluatie is in dit onderzoek geen sinecure gebleken. Om te komen tot een methodiek voor het bepalen van de kwaliteit van één beleidsvaluatie op zich is blijkt niet het probleem te zijn, maar een methodiek ontwikkelen die in verschillende situaties toepasbaar is, is een stuk moeilijker. Toch is daartoe een poging gedaan.

In de praktijk blijkt dat veel evaluatieonderzoek niet goed wordt uitgevoerd. De belangrijkste oorzaak hiervoor is het ontbreken van inzicht in het beleidsveld. Hierdoor kunnen relaties tussen beleidsmaatregelen en het optreden van effecten niet goed in kaart worden gebracht. Dit heeft zijn uitwerking op de gehele uitvoering van een beleidsvaluatie. Het optreden van effecten en het verklaren ervan kan nooit duidelijk worden gemaakt. Daarnaast spelen nog andere oorzaken een rol, zoals gebrekkige doelstellingen en het veranderen van het beleid tijdens de uitvoering van de evaluatie.

Opvallend is dat in enkele geëvalueerde evaluatieonderzoeken de titel niet de lading van het onderzoek dekt. Dit wordt vooral veroorzaakt door het feit dat onduidelijkheid bestaat over wat de begrippen monitoring, auditing en evaluatie inhouden. Hierdoor kan verwarring ontstaan bij zowel de opdrachtgever als de onderzoeker.

Tot slot is gebleken dat de in dit onderzoek opgestelde methodiek voor het uitvoeren van meta-evaluatie niet compleet is, maar wel een goed hulpmiddel kan zijn bij snel inzichtelijk maken van de kwaliteit van een evaluatieonderzoek. Daarnaast is het van belang om te onderkennen dat er een wisselwerking bestaat tussen het beleid en de evaluatie ervan. Veel winst kan nog worden geboekt in de efficiëntie van zowel het beleid als de evaluatie wanneer tijdens het opstellen van beleid rekening wordt gehouden de evalueerbaarheid ervan.





## Voorwoord

Het voorliggende onderzoeksrapport is mijn afstudeerscriptie voor de studie Technische Planologie aan de Rijksuniversiteit Groningen. Deze scriptie is het voortvloeiende uit mijn stage die ik maanden geleden gelopen heb bij Grontmij Drenthe in Assen. Een onderdeel van deze stage was de opdracht voor het uitvoeren van een “Beleidseffectrapportage Goederenvervoer Noord-Nederland”. Dit zou in eerste instantie ook mijn afstudeerscriptie worden. Tijdens het project begon bij mij echter steeds meer het besef door te dringen dat het uitvoeren van zo’n beleidseffectrapportage veel meer haken en ogen kent dan op het eerste gezicht lijkt. Vanuit mijn wetenschappelijke achtergrond ben ik zeer ambitieus begonnen aan het project, maar naar mate de tijd verstreek merkte ik dat de praktijk niet altijd te combineren viel met het geleerde aan de universiteit. Vooral het tijd- en kostenaspect blijkt bij een commercieel adviesbureau een belangrijke, van invloed zijnde, factor. Aan de ene kant werkte dit frustrerend, aan de andere kant was dit een interessante constatering. Hierdoor werd wel mijn nieuwsgierigheid geprikkeld naar het antwoord op de vraag of het uitvoeren van een correct evaluatieonderzoek überhaupt mogelijk zou zijn. Het antwoord zal ik hier niet geven, anders heeft het lezen van dit rapport weinig zin.....

Vanaf deze plek wil ik iedereen bedanken voor zijn of haar bijdrage aan het realiseren van deze scriptie. Speciale dank gaat echter uit naar Grontmij Drenthe, en Hans Popken in het bijzonder. Zonder zijn ondersteuning, enthousiasme en begrip voor het soms moeizaam verlopen van het project had ik dit voorwoord niet eens hoeven schrijven. Daarnaast wil ik Mirjan Bouwman van de Rijksuniversiteit Groningen bedanken voor haar begeleiding en eindeloze geduld met het vorderen van mijn scriptie, en de altijd interessante gesprekken en opbouwende kritiek van haar kant.

Tot slot: pa en ma, bedankt voor het onvoorwaardelijke vertrouwen en de jarenlange steun om mijn studie tot een goed einde te brengen. En Hanneke: bedankt voor de “schoppen onder mijn kont” de afgelopen jaren en het enorme geduld dat je hebt opgebracht om mij bij het moeizame proces tot het afronden van mijn scriptie terzijde te staan.

Groningen, september 2004

Erik Spoelstra



# INHOUDSOPGAVE

<b>SAMENVATTING .....</b>	<b>I</b>
<b>VOORWOORD .....</b>	<b>VI</b>
<b>INHOUDSOPGAVE .....</b>	<b>1</b>
<b>1 INLEIDING .....</b>	<b>3</b>
1.1 INLEIDING .....	3
1.2 VRAAGSTELLING EN DOELSTELLING .....	4
1.3 LEESWIJZER.....	6
<b>DEEL I - THEORETISCH KADER.....</b>	<b>7</b>
<b>2 BELEID EN EVALUATIE.....</b>	<b>7</b>
2.1 INLEIDING .....	7
2.2 BELEID.....	7
2.3 EISEN AAN BELEID .....	9
2.4 BELEIDSONDERZOEK EN -EVALUATIE.....	10
2.5 ONDERZOEKSTYPEN.....	12
<b>3 UITVOERING VAN EVALUATIEONDERZOEK.....</b>	<b>17</b>
3.1 INLEIDING .....	17
3.2 VOORBEREIDINGSFASE.....	17
3.3 UITVOERINGSFASE.....	18
3.4 AFRONDINGSFASE.....	22
3.5 VALKUILEN EN PROBLEMEN BIJ EVALUATIEONDERZOEK .....	23
<b>4 META-EVALUATIE.....</b>	<b>27</b>
4.1 INLEIDING .....	27
4.2 META-EVALUATIE .....	27
4.3 META-EVALUATIE-CHECKLISTS.....	28
<b>5 METHODE VAN ONDERZOEK .....</b>	<b>33</b>
<b>DEEL II – DE PRAKTIJK .....</b>	<b>39</b>
<b>6 CASESTUDIES.....</b>	<b>39</b>
6.1 INLEIDING .....	39
6.2 BELEIDSEFFECTRAPPORTAGE VERKEER & VERVOER NOORD-NEDERLAND, PERIODE 1986-2001 .....	40
6.3 BELEIDSEFFECTRAPPORTAGE VERKEER EN VERVOER 1999.....	41
6.4 INTERPROVINCIALE RAPPORTAGE MILIEU, WATER, LANDBOUW EN NATUUR 2002.....	43
6.5 EVALUATIE VAN DE KAMPEERREGELGEVING – 2002 .....	45
6.6 EVALUATIEONDERZOEK WET- EN REGELGEVING MILIEUJAARVERSLAGEN .....	47
6.7 MONITORING EN EVALUATIE DEREGULERING TAXIVERVOER 1999-2001 .....	49
6.8 EVALUATIE VAN DE DERDE KAMER .....	52
6.9 CONCLUSIE BEOORDELING EVALUATIEONDERZOEKEN .....	54
<b>7 CONCLUSIE EN AANBEVELINGEN .....</b>	<b>57</b>
7.1 INLEIDING .....	57
7.2 CONCLUSIES.....	57
7.3 AANBEVELINGEN .....	59
7.4 TOT SLOT .....	60
<b>LITERATUURLIJST .....</b>	<b>63</b>
<b>BIJLAGEN .....</b>	<b>67</b>





# 1 Inleiding

## 1.1 Inleiding

*“De overheid faalt op tal van fronten, wat leidt tot ergernis en leedvermaak bij de wantrouwende burger. [...] Is er wel een overheidsdienst die functioneert?” – De Volkskrant, 28 december 2002.*

*“Nederland slaagt er met geen mogelijkheid in te voldoen aan zijn nationale en Europese doelstellingen voor de toepassing van duurzame energie, blijkt uit onderzoek.[...] Het kabinetsbeleid op het gebied van duurzame energie dreigt een faliekante mislukking te worden. Uit een aantal recente onderzoeken blijkt dat Nederland ruimschoots tekortschiet bij het realiseren van de nationale en Europese doelstellingen voor de toepassing van duurzame energie - zelfs niet als aan alle bestaande afspraken en beleidsvoornemens wordt voldaan.[...] Het beleid is te vrijblijvend. Het ontbreekt in Den Haag aan daadkracht en visie.” – Het Financiële Dagblad, 19 juni 2004.*

*“Het beleid van de overheid tegen de verdroging van de natuur faalt.” - Het Algemeen Dagblad, 22 mei 2004.*

*“Alles bij elkaar is het IJsselmeer het meest overbeviste water ter wereld, chargeert de ecooloog. 'Slechts een op de duizend palingen krijgt kans te ontsnappen om zich voort te planten. De ecologisch schadelijke bijvangst van eenjarige baars en snoekbaars in de fuiken is enorm. En we kunnen dat de beroepsvissers niet eens kwalijk nemen. De overheid reguleert dit met vergunningen. Het is dus het overheidsbeleid dat faalt.’” – De Volkskrant, 27 maart 2004.*

*“Er is een groot tekort aan geschikte woningen en zorg voor ouderen die zelfstandig willen blijven wonen. Bovendien heeft de overheid geen idee hoe groot de vraag van ouderen hiernaar precies is. Het beleid om meer ouderen in staat te stellen zelfstandig te laten wonen, heeft gefaald.” - NRC Handelsblad, 6 juni 2003.*

*“De overheid is er de afgelopen jaren veel te weinig in geslaagd mensen met een uitkering aan het werk te krijgen.[...] Wel constateert het SCP dat het beleid om het aantal WAO'ers terug te dringen de afgelopen jaren is mislukt, ondanks het tot voor kort gunstige economische klimaat en de vele vacatures.” - Het Financiële Dagblad, 14 september 2002.*

*“Het rapport “Tussen beleid en uitvoering” legde de gegroeide kloof tussen de almaar toenemende hoeveelheid wetten en regels en de uitvoering en handhaving ervan bloot. Het bleek op geen van de 29 door de Rekenkamer onderzochte terreinen zeker dat het beleid volledig en goed is uitgevoerd. Of informatie om dit vast te stellen ontbreekt. Of het is duidelijk dat beleid niet helemaal of niet goed genoeg is uitgevoerd. Er is slechte aansluiting tussen beleid en uitvoering. Problemen die het beleid geacht wordt op te lossen blijven daardoor vaak langer bestaan dan in de besluitvorming bedoeld was, concludeerde de Rekenkamer.” - NRC Handelsblad, 13 juni 2003.*

Zomaar een greep uit een aantal krantenberichten van de afgelopen tijd. De overheid doet haar best om beleid op te stellen om bepaalde maatschappelijke problemen op te lossen. Vaak lijkt zij echter niet te slagen in haar opzet. Veel beleid dat door de jaren heen is opgesteld lijkt niet te werken. En dit geldt niet voor één beleidsterrein, maar op tal van beleidsgebieden. Natuurlijk zijn ook voorbeelden te noemen die bewijzen dat beleid wel

degelijk succesvol kan zijn. Maar hoe kan het dat één beleid wel succesvol is en het andere niet? Duidelijk is dat er geen standaardmodel bestaat voor het opstellen van beleid, want als dat zou bestaan dan zou beleid toch altijd succesvol moeten zijn... of toch niet? Blijkbaar spelen zoveel factoren een rol dat het opstellen van succesvol beleid een complexe aangelegenheid is.

Maar wat zijn dan toch de redenen voor het falen van veel beleid? Als beleid zo vaak faalt, dan zouden toch ook redenen aan te wijzen moeten zijn waaruit lering kan worden getrokken. Maar is dit zo en, zo ja, wordt er daadwerkelijk lering getrokken uit gemaakte fouten? Dit zijn allemaal vragen die zich doen gelden bij het onderzoek naar het succes en falen van beleid, oftewel *beleidsonderzoek*.

## 1.2 Vraagstelling en doelstelling

Dit voorliggende onderzoek richt zich echter niet rechtstreeks op deze vragen, maar zoekt meer naar een antwoord op een vervolgvraag: Hoe kan men antwoorden vinden op deze genoemde vragen? Dus, hoe moet men een onderzoek inrichten om antwoord te krijgen op de vragen die zich bij beleidsonderzoek voordoen. Beleidsevaluatie is een vorm van beleidsonderzoek waarin in het functioneren van het beleid centraal staat. Hierbij wordt, indien beleidsevaluatie goed wordt uitgevoerd, gekeken naar de het succes of falen van het beleid én de oorzaken daarvan. Maar wordt beleidsevaluatie wel altijd goed uitgevoerd?

Beleidsevaluatie wordt in dit onderzoek centraal gesteld, beleidsevaluatie wordt zelf onderwerp van onderzoek. Beleidsevaluatieonderzoek richt zich op de kwaliteit van de beleidsevaluaties. Wat opvalt is dat eigenlijk iedereen wel een mening over het uitgevoerde of uit te voeren beleid heeft. Dit is op zichzelf ook een vorm van beleidsevaluatie, iedereen bepaalt aan de hand van door zichzelf opgestelde criteria de waarde van het beleid, voor zichzelf. Maar worden er ook wetenschappelijke eisen gesteld die bepalen of een beleidsevaluatie goed is uitgevoerd? De krantenknipsels duiden op een vorm van beleidsevaluatie, maar het ontbreekt ons aan inzicht in de uitgevoerde beleidsevaluaties en de gehanteerde methodiek. Dit is wel nodig om de geldigheid van de gedane uitspraken te kunnen bepalen.

In dit onderzoek staat dus niet zozeer het succes of falen van een beleid centraal (en de redenen hiervoor), als wel de evaluatie van beleid. De wijze waarop dit gebeurt laat namelijk vaak te wensen over. Conclusies worden getrokken op basis van niet voldoende onderbouwde argumenten of een te magere 'bewijslast'. Toch wordt aan veel beleidsevaluaties en zogenaamde beleidseffectrapportages veel waarde gehecht (en ook veel tijd en geld besteedt). In de jaren zestig waren planners vaak nog heel optimistisch aangaande de mogelijkheden om zowel efficiëntie als effectiviteit van beleid te vergroten op basis van wetenschappelijk onderzoek. De praktijk heeft echter geleerd hoe moeizaam het proces is van specificeren (nader concretiseren) en het implementeren (daadwerkelijk tot uitvoering doen komen) van beleid. De onderkenning van het belang van beleidsevaluatie is in Nederland sinds de jaren zeventig toegenomen. Omdat het overheidsbeleid vaak opgesteld wordt voor de voor gehele samenleving, is de uitwerking van het beleid dus ook van belang voor de gehele samenleving. Beslissingen die worden genomen op basis van foutieve aannames als gevolg van verkeerd uitgevoerde beleidsevaluatieonderzoeken moeten zoveel mogelijk worden voorkomen, omdat ze grote gevolgen kunnen hebben voor de maatschappij.

Tegenwoordig is beleidsevaluatie een populair instrument om de effectiviteit, efficiëntie en aanvaardbaarheid van overheidsbeleid te meten. Een van de belangrijkste verklaringen voor de grote belangstelling voor de uitvoering van overheidsbeleid, is het groeiende besef dat de effecten van een beleid in sterke mate worden bepaald door het verloop van de beleidsuitvoering. Evaluatie kan in het proces van beleidsuitvoering een belangrijke rol spelen.

Het spreken over succes en falen van de beleidsuitvoering veronderstelt maatstaven waaraan het succes of falen wordt afgemeten (Hoogerwerf, 1998). In de praktijk blijkt de methodiek nogal eens te rammelen. Om te bepalen of een beleidsevaluatie goed is uitgevoerd zijn verschillende methoden in omloop, onder de noemer *meta-evaluatie*. Algemeen geldt dat wanneer een evaluatie zelf onderwerp is van een evaluatieonderzoek dit meta-evaluatie wordt genoemd. In de Verenigde Staten heeft men meer ervaring met het uitvoeren van meta-evaluaties, hoewel in Nederland dit ook steeds vaker gebeurt.

In dit onderzoek wordt een meta-evaluatie uitgevoerd van enkele (overheids)beleidsevaluatieonderzoeken. De reden voor het uitvoeren van beleidsevaluatieonderzoek is in de meeste gevallen inzicht willen krijgen in het 'presteren' van het beleid. Aan deze onderzoeken wordt door een aantal beleidsmakers, media, maatschappij, enzovoorts, in veel gevallen een groot belang gehecht en worden daarom bepaalde beslissingen genomen die gebaseerd zijn op het beleidsevaluatieonderzoek. De vraag is of dit altijd even terecht is. Dat is de eerste reden voor het uitvoeren van dit onderzoek.

Maar er is nog een tweede reden, namelijk de behoefte aan een methodiek die inzicht kan geven in de kwaliteit van beleidsevaluatieonderzoek, die in voor verschillende situaties toepasbaar is. Hieruit vloeit voort dat, door de verwachte wisselwerking tussen beleid en beleidsevaluatie, de kwaliteit van beleid ook toe zou kunnen nemen. Dit is de derde reden voor het uitvoeren van dit onderzoek.

De voor dit onderzoek geformuleerde algemene vraagstelling luidt dan als volgt:

*Waar ligt de kwetsbaarheid van beleidsevaluatieonderzoek en hoe kan er een methode worden ontwikkeld om beleidsevaluatieonderzoek op een efficiënte manier te kunnen beoordelen en te verbeteren en kan hiermee ook het opstellen van beleid worden verbeterd?*

Het hieruit voortvloeiende doel van onderzoek is drieledig:

1. Het in kaart brengen van de kwetsbaarheid van beleidsevaluatieonderzoek;
2. Het ontwikkelen van een methodiek voor het bepalen van de kwaliteit van beleidsevaluatieonderzoek;
3. Het toepasbaar maken van de methode voor het verbeteren van beleidsevaluatieonderzoek én beleid.

Het hoofddoel is dus enerzijds het in kaart brengen van de kwaliteit van beleidsevaluatieonderzoeken, anderzijds om de voorkomende tekortkomingen zoveel mogelijk te beperken en op te lossen, door een methode aan te reiken voor waarmee de kwaliteit van evaluatieonderzoek kan worden bepaald.

Hiertoe is de volgende werkwijze gehanteerd. Allereerst zijn de begrippen die van belang zijn voor het onderzoek nader toegelicht. Het is van belang om inzicht te hebben in de betekenis van de begrippen, omdat deze door het hele onderzoek een centrale rol spelen. Daarnaast verduidelijkt de uiteenzetting van de verschillende begrippen de onderlinge verhoudingen en relaties.

Om tot een methode te komen voor het bepalen van de kwaliteit van beleidsevaluatieonderzoek, oftewel het uitvoeren van een meta-evaluatie, zijn methoden uit de literatuur en praktijk verzameld en samengevoegd. Daarnaast zijn enkele evaluatieonderzoeken die de afgelopen jaren zijn uitgevoerd geïnventariseerd en beoordeeld aan de hand van de ontwikkelde methode. Hierbij bestaat niet de illusie volledig te zijn, aangezien het ondoenlijk is om alle beleidsevaluatieonderzoeken van de afgelopen jaren te evalueren. Het hoofddoel van dit onderzoek is om een methodiek aan te reiken, waarmee de

kwaliteit van evaluatieonderzoek kan worden bepaald. Dit onderzoek is daarom beperkt tot enkele voorbeelden uit de ruimtelijke ordening, om enig inzicht te krijgen in de beperkingen die voorkomen bij evaluatieonderzoek. Een kritische kanttekening is echter wel op zijn plaats, namelijk het feit dat men al decennia lang bezig is om een goede methodiek te ontwikkelen voor meta-evaluatie geeft aan dat dit een zeer complexe aangelegenheid is. Een methodiek is nooit volledig, dus de in dit onderzoek opgestelde en gehanteerde ook niet.

Tot slot worden de conclusies van dit onderzoek vermeld en worden aanbeveling gedaan voor het uitvoeren van evaluatieonderzoek.

### **1.3 Leeswijzer**

Dit onderzoek bestaat uit twee delen: deel 1, waarin het theoretisch kader centraal staat, en deel 2, waarin het uitgevoerde onderzoek in de praktijk wordt toegelicht. Deel 1 bevat vier hoofdstukken, namelijk hoofdstuk 2 tot en met 5, deel 2 bevat de hoofdstukken 6 en 7.

In deel 1 wordt allereerst in hoofdstuk 2 de begrippen 'beleid' en 'evaluatie' uitgewerkt. Hierdoor wordt inzicht verkregen in de deze twee sleutelbegrippen en hun onderlinge verhouding. In hoofdstuk 3 wordt de uitvoering evaluatieonderzoek toegelicht. Dit gebeurt aan de hand van methoden zoals die in de literatuur terug zijn te vinden. In hoofdstuk 4 staat een specifiek onderdeel van evaluatieonderzoek centraal, namelijk meta-evaluatie. Hierin wordt duidelijk gemaakt hoe evaluatieonderzoek zelf onderwerp van een evaluatie kan worden gemaakt. In hoofdstuk 5 tenslotte wordt de methode van onderzoek opgesteld en onderbouwd.

In het eerste hoofdstuk van deel 2, hoofdstuk 6 wordt de methode van onderzoek toegepast op de beleidsevaluaties uit de praktijk en worden per studie conclusies getrokken. In hoofdstuk 7 worden deze conclusies naast elkaar gelegd en wordt gekeken of er verbanden te ontdekken zijn en zogenaamde 'algemene lessen' zijn te trekken. Daarnaast worden in hoofdstuk 7 aanbevelingen gedaan om het onderzoek naar beleidsevaluatie te verbeteren, en ook om de kwaliteit van beleidsevaluatie op zich te verbeteren.

## DEEL I - THEORETISCH KADER

### 2 Beleid en evaluatie

#### 2.1 Inleiding

In dit hoofdstuk wordt ingegaan op de theorie rondom evaluatie en beleidsevaluatie in het bijzonder. De uitgebreide uiteenzetting van begrippen, methoden en problemen die bij het uitvoeren van evaluatie voorkomen, wordt voor het onderzoek van belang geacht, omdat het inzicht geeft in de complexiteit die met wetenschappelijk evaluatieonderzoek samenhangt. In dit hoofdstuk is de verhouding tussen evaluatie en beleid uitgewerkt, en worden methoden voor (beleids)evaluatie weergegeven, evenals de bijbehorende problemen en valkuilen.

In paragraaf 2.2. wordt kort ingegaan op beleid in algemene zin en de effecten ervan. Vervolgens worden in paragraaf 2.3. de begrippen 'evaluatie' en 'beleidsevaluatie' uitgewerkt. In paragraaf 2.4 worden de eisen aan beleid behandeld. In paragraaf 2.5. komen de verschillende onderzoekstypen aan de orde.

#### 2.2 Beleid

Beleid in algemene zin wordt door Voogd (1995a) gedefinieerd als "een complex van handelingen ten aanzien van een probleem of doelgroep, (meestal) vastgelegd in plannen. [...] Het bestaat uit een veelheid aan componenten: doelstellingen, middelen, tijdsaanduidingen, instrumenten, activiteitschema's en randvoorwaarden". Dit is een uitgebreide opsomming van componenten vergeleken met bijvoorbeeld Blommestein (1984), die beleid ziet als een min of meer overwogen streven bepaalde doeleinden met bepaalde middelen in een bepaalde tijdsvolgorde te bereiken. Herweijer en Hoogerwerf (1998) zijn het met deze definitie van beleid grotendeels eens. Zij geven een opsomming van uit de literatuur verzamelde definities voor beleid, die gemeen hebben dat ze de drie sleutelbegrippen "doeleinden", "middelen" en "tijdsvolgorde of -keuze" bevatten. Deze begrippen houden dus in dat beleid bestaat uit één of meer doelstellingen die met een bepaalde inspanning binnen een bepaalde periode gehaald dienen te worden. Daarnaast voegen zij hier aan toe dat beleid een poging tot een antwoord op een probleem moet inhouden.

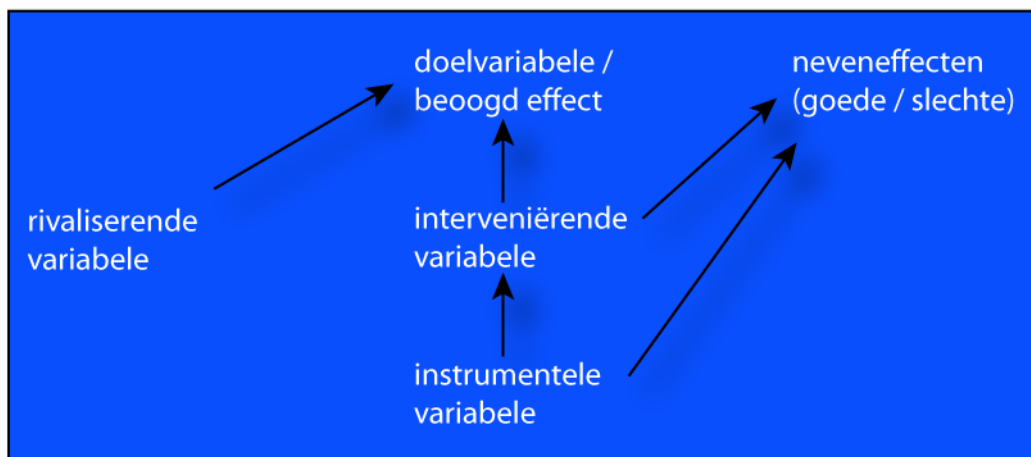
Een kenmerk van beleid is dat iedereen het maakt, ongeacht of men zich daar bewust van is of niet. Niet iedereen heeft echter dezelfde doeleinden of doelstellingen met hun, zijn of haar beleid. Er bestaat een belangrijk verschil tussen overheidsbeleid en beleid dat door één of meer individuen wordt opgesteld. Aangezien de centrale functie van de overheid volgens Herweijer en Hoogerwerf (1998) 'het regeren', dat wil zeggen het totstandbrengen, voeren en uitvoeren van een beleid *voor en namens gehele samenleving* is, kan overheidsbeleid gezien worden aan een poging tot het oplossen of het verminderen van een maatschappelijk probleem. Een beleid van een individu (of groep individuen) is meer gericht op het behalen van eigen voordeel of oplossen van eigen problemen. Een gevolg hiervan is dat overheidsbeleid en individueel beleid (of van een groep individuen) tegenstrijdig kunnen zijn en elkaar kunnen tegenwerken.

Een ander kenmerk van beleid is volgens de Vries (2001) dat het dynamisch van aard is. Om dit te onderschrijven citeert hij Ringeling (1978): "Beleid moet niet worden opgevat als een schema van wenselijkheden die voor het gehele beleidsproces blijven gelden. Beleid wordt

continu geformuleerd en gewijzigd tijdens de uitvoeringsfase. Middelen, tijdstippen en zelfs doelstellingen worden veranderd terwijl het beleid wordt gevoerd”<sup>1</sup>.

Het kenmerk van beleid dat centraal staat in dit onderzoek is echter dat het gevolgen heeft voor het onderwerp waarop het gericht is. Dit worden de effecten van beleid genoemd. In algemene zin kunnen effecten worden getypeerd als een verandering in een bestaande situatie die wordt veroorzaakt door een gebeurtenis of een handeling (Ministerie van Financiën, 1993). De constatering dat beleid gevolgen heeft voor het onderwerp waarop het gericht is, betekent dat overheidsbeleid effect heeft op de gehele maatschappij. Overheidsbeleid is immers een poging tot het oplossen van een maatschappelijk probleem.

Beleids effecten kunnen in verschillende categorieën worden ingedeeld. Zo zijn er de effecten die de bedoeling van de beleidsmakers zijn en die als doelstellingen worden vastgelegd in het beleid. Deze zogenaamde beoogde effecten zijn de effecten die met behulp van een maatregel of ingreep doelbewust teweeg worden gebracht. Het kan echter ook voorkomen dat effecten optreden die niet voorzien en/of bedoeld waren (ook op onderwerpen waarop het beleid niet gericht is). Dit zijn de zogenaamde niet-beoogde effecten, of neveneffecten (Ministerie van Financiën, 1993). In Figuur 2.1 wordt dit verduidelijkt.



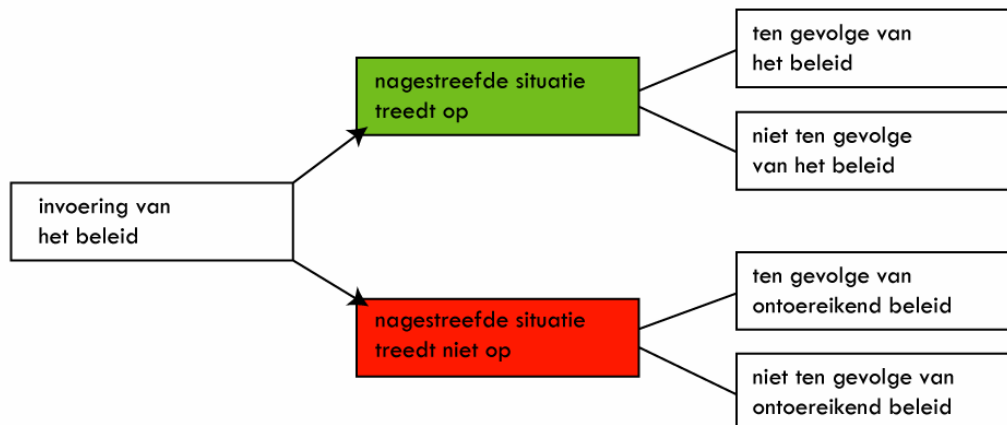
**Figuur 2.1 Relatie tussen variabelen en effecten (Ministerie van Financiën, 1998)**

Uit Figuur 2.1 blijkt dat het optreden van effecten van meerdere factoren afhankelijk is en dat beleid niet volledig verantwoordelijk kan zijn voor het al dan niet optreden van effecten. Het beleid is het uitgangspunt, maar het beleid kan versterkt of tegengewerkt worden door de intervariërende variabele (externe invloed), waardoor de invloed van het beleid op het beoogde effect (doelstelling) minder groot of juist groter is dan zonder externe invloed. Ook kunnen door zowel de instrumentele variabele als de intervariërende variabele neveneffecten optreden (positief of negatief).

Een ander onderscheid in effecten is dat in directe en indirecte effecten. Effecten die rechtstreeks voortvloeien uit de invoering van een beleid worden directe effecten genoemd. Indirecte effecten zijn in dit geval doorwerkingen; zij vloeien voort uit de directe. Deze indirecte effecten worden regelmatig over het hoofd gezien, terwijl zij toch ook mede bepalend zijn voor het succes of falen van een beleid.

<sup>1</sup> Dit is een belangrijke opmerking in het kader van dit onderzoek: de Vries zegt hiermee indirect dat een beleidsevaluatieonderzoek met betrekking tot beleid moeilijk is uit te voeren, aangezien de randvoorwaarden continu wijzigen.

De relatie tussen beleid en het optreden van effecten heeft vier mogelijkheden tot gevolg. Deze mogelijkheden worden in Figuur 2.2 weergegeven.



**Figuur 2.2** Beleid en optreden effecten

In Figuur 2.2 wordt verduidelijkt dat de kans bestaat dat de onderzoeker tot de conclusie komt dat een beleid een bepaald effect tot gevolg heeft gehad, maar dat hij geen rekening houdt dat de doelbereiking eventueel ook door andere factoren kan worden versterkt en vice versa. Directe effecten zijn niet altijd van grotere betekenis dan indirecte effecten. Integendeel, volgens onder andere het Ministerie van Financiën (1993a) gaat het juist in veel gevallen om doorwerkingen van beleidsmaatregelen. Figuur 2.2 sluit aan bij Figuur 2.1 waaruit ook blijkt dat andere factoren een rol kunnen spelen bij het bereiken van de doelstellingen. In de praktijk blijkt het vaak voor te komen dat een onderzoeker de conclusie trekt dat een beleid succesvol is (of juist niet) omdat de doelstellingen zijn bereikt, zonder daarbij de invloed van het beleid daarbij te betrekken. Het onderzoek dat de naam evaluatie- of effectiviteitsonderzoek draagt, beperkt zich dan tot de “tweede” stap in de figuur: vaststellen van het al dan niet optreden van een gewenste situatie, zonder dat de bijdrage van het beleid hierbij wordt vastgesteld<sup>2</sup>.

### 2.3 Eisen aan beleid

Aangezien beleid gezien wordt als een min of meer weloverwogen streven bepaalde effecten teweeg te brengen, kunnen met betrekking tot deze effecten eisen aan het beleid worden gesteld. Als eerste eis kan gesteld worden dat beleid een bepaalde mate van doelbereiking moet hebben. Dit is immers de eerste aanleiding voor het opstellen van beleid: men wil een bepaald doel bereiken. Daarnaast komt Hoogerwerf (1998), na bestudering van de literatuur met betrekking tot beleidsonderzoek, tot twee eisen aan beleid die betrekking hebben op de doelbereiking, en nog een stap verder gaan, namelijk: effectiviteit en efficiëntie<sup>3</sup>. Deze begrippen hebben de volgende betekenis:

*Effectiviteit* is de mate waarin beleid bijdraagt aan het bereiken van de in het beleid vastgelegde doelstellingen, die opgevat kunnen worden als beoogde effecten. Oftewel: de effectiviteit van een beleid is de mate van doelbereiking voor zover deze aan het betreffende beleid is te danken. Bij deze formulering zijn de relatie en het verschil met het begrip doelbereiking duidelijk. Bressers (1984) geeft dit als volgt aan:

<sup>2</sup> Dit komt overeen met de tweede gradatie die Arts (1998) aanbrengt in evaluatieonderzoek, namelijk auditing zie paragraaf 2.5.

<sup>3</sup> Wanneer niet aan de eis van doelbereiking wordt voldaan, heeft het geen zin om te spreken over effectiviteit of efficiëntie. Hier wordt later in dit hoofdstuk verder op ingegaan.



*“Enerzijds is niet alle doelbereiking ook effectiviteit van het beleid te noemen. Allerlei andere factoren naast het beleid kunnen elementen waarover doeleinden zijn geformuleerd in de gewenste richting beïnvloeden. In dat geval is de effectiviteit van het beleid kleiner dan de doelbereiking. Anderzijds kan ook het omgekeerde zich voordoen. Allerlei factoren kunnen dermate tegenzitten dat er zelfs geen sprake van enige verbetering hoeft te zijn om toch te kunnen concluderen tot een grote effectiviteit van het beleid wegens het voorkomen van erger”. (p.83)*

Effectiviteit wordt ook wel omschreven als de doeltreffendheid van beleid.

*Efficiëntie* houdt de doelmatigheid van beleid in, met andere woorden de verhouding tussen de ‘kosten en baten’ van een beleid. Hierbij gaat het om de inspanning die geleverd moet worden om een doelstelling te bereiken.

Hoogerwerf (1998) stelt nog een eis aan beleid, die vaak over het hoofd wordt gezien, namelijk *legitimiteit*. Legitimiteit is de mate waarin het beleid door de betrokken partijen gedragen wordt, oftewel als aanvaardbaar wordt beschouwd. Dit heeft niet zozeer met de ‘afrekening’ van het beleid op de daarin geformuleerde doelstellingen te maken, als wel met het draagvlak onder betrokkenen partijen (in beleid wordt niet vaak gesproken over het draagvlak onder betrokkenen in termen van harde doelstellingen). De Vries (2001) meent dat wanneer in een democratisch bestel een meerderheid van burgers vindt dat de overheid tekortschiet (ook al is daar geen ‘hard’ bewijs voor), dit consequenties heeft voor de legitimiteit van het overheidshandelen. Hij meent dan ook dat het van groot belang is om inzicht te krijgen in de mate waarin de overheid in werkelijkheid presteert en hoe dit beoordeeld kan worden. Tevens dient de doeltreffendheid van het overheidsbeleid hiermee te verbeteren. Indien overheidsbeleid niet wordt gedragen door de betrokkenen, is het bij voorbaat vrijwel kansloos. Toch zal de overheid ook beleid moeten doorvoeren (ten behoeve van het algemeen belang) dat bij bepaalde individuen of groepen op weinig steun kan rekenen. Deze individuen of groepen zullen proberen de invoering van dit beleid zoveel mogelijk te frustreren. De overheid kan in dit geval instrumenten inzetten om toch het beleid door te voeren, maar de efficiëntie van het beleid neemt daardoor af. Het ‘kost’ immers meer om het beleid te implementeren en tot uitvoering te brengen.

Aan de drie genoemde voorwaarden die volgens Hoogerwerf (1998) aan beleid kunnen worden gesteld, wordt volgens hem door de overheid lang niet altijd voldaan. Dit kan verschillende oorzaken hebben, maar hij ziet als belangrijkste het ontbreken van winststreven en prikkels van concurrentie, en het bureaucratische karakter van de overheid.

Wat opvallend is dat Hoogerwerf de doelbereiking niet als eis aan beleid stelt, terwijl in veel literatuur weer de eis van legitimiteit ontbreekt. Doelbereiking wordt ook in dit onderzoek niet als harde eis aan beleid gesteld, aangezien de doelen ook bereikt hadden kunnen worden zonder de invoering van beleid. Doelbereiking kan dus niet als eis worden gesteld, omdat het min of meer los moet worden gezien van het beleid (zie Figuur 2.2). Hoogerwerf kiest daarom ook voor effectiviteit van beleid als eis, doordat de invloed van beleid daarin wel wordt opgenomen. In dit onderzoek wordt uitgegaan van drie eisen aan beleid: doeltreffendheid, doelmatigheid en legitimiteit. Het onderzoek naar de mate waarin beleid voldoet aan deze eisen maakt deel uit van (overheids)beleidsonderzoek. Hierop wordt in de volgende paragraaf nader ingegaan.

#### **2.4 Beleidsonderzoek en -evaluatie**

Zoals gezegd heeft overheidsbeleid gevolgen voor de gehele maatschappij. Daarom is het van belang om inzicht te krijgen in deze effecten en ongewenste ontwikkelingen als gevolg van het beleid zoveel mogelijk uit te sluiten. De behoefte aan evaluatie van beleid is vooral sinds het einde van de jaren vijftig gegroeid, maar het evaluatieonderzoek gericht op

overheidsbeleid, is pas in de jaren tachtig in Nederland sterk toegenomen<sup>4</sup> (Bressers en Hoogerwerf, 1991). Deze toegenomen belangstelling valt vooral te verklaren door het besef dat veel overheidsbeleid onvoldoende doeltreffend is, onvoldoende doelmatig is en onvoldoende door de betrokken burgers wordt gesteund (effectiviteit, efficiëntie en legitimiteit). De overheid beschikt daarnaast vaak amper over voldoende, of voldoende precieze informatie om een beleid op te baseren. Dit heeft onder andere tot gevolg gehad dat in nieuwe wetten werd voorgeschreven dat de werking van de een wet of beleid na verloop van aantal jaren moet worden geëvalueerd, om zo evaluatie en dataverzameling in een vroeg stadium te verankeren.

Om het succes of falen van beleid te kunnen bepalen, wordt er beleidsonderzoek uitgevoerd. Het begrip beleidsonderzoek is in 1980 door van Vall geïntroduceerd (van Hoesel, 2003). Beleidsonderzoek kan volgens van Hoesel (2003) worden opgevat als “een verzamelterm waaronder diverse onderzoeksactiviteiten worden begrepen die ten behoeve van de beleidsontwikkeling worden uitgevoerd, zoals voorspellingen van planbureaus, *beleidsevaluaties* van inspecties, probleemverkenningen en beleidanalyses door universiteiten, grootschalige publieksenquêtes en casestudies door zelfstandige onderzoeksbureaus.” (p.3). Beleidsonderzoek heeft dus ten doel de beleidsontwikkeling te verbeteren, evaluatieonderzoek vormt hier een belangrijk onderdeel van, met name met het oog op terugkoppeling. Van gemaakte fouten kan men leren. Evaluatieonderzoek kan volgens van Hoesel (2003) worden gezien als de belangrijkste deelverzameling van beleidsonderzoek, omdat het gericht is op verwachte of daadwerkelijke effecten van beleid en ook op de storende factoren die het beoogde effect in de weg staan. Hij voegt nog toe dat beleidsevaluatieonderzoek altijd beleidsonderzoek is, maar andersom gaat dit niet op. Evaluatieonderzoek wordt door enkelen echter zo ruim opgevat dat het vrijwel synoniem wordt met beleidsonderzoek.

Voor dit onderzoek is het van belang om aan te geven dat de nadruk niet ligt op beleidsonderzoek, maar op evaluatieonderzoek. Evaluatieonderzoek maakt echter wel deel uit van beleidsonderzoek zodat dit onderzoek ook voor een deel als beleidsonderzoek gezien kan worden.

Volgens Bressers en Hoogerwerf (1991) kan evaluatie in algemene zin gezien worden als het beoordelen van de voorstelling of waarneming van een bepaald verschijnsel (een bepaalde situatie of verandering daarin) aan de hand van bepaalde criteria. Douglass (1998) is het hier in grote mate mee eens, maar voegt daar nog aan toe dat evaluatie ook als doel moet hebben een project of beleid te verbeteren of bij te sturen. Volgens Douglass is het onmogelijk om één juiste definitie van evaluatie te geven. Wel geeft hij een interessante en duidelijke definitie weer van Steele (1991) die als volgt luidt:

*“Evaluation is the act of comparing what should be (criteria, standards, goals or objectives) with that what is (evidence, data, information) for the purpose of ascertaining the worth or value of what is being evaluated”.*

Simpeler is evaluatie eigenlijk niet te omschrijven; het vergelijken van ‘wat zou moeten zijn’ met dat ‘wat is’, aan de hand van data over criteria, met als doel de waarde van wat geëvalueerd is te bepalen.

Ondanks dat, zoals ook Douglass aangaf, het onmogelijk is om één juiste definitie voor evaluatie te geven, plaatsen Bressers en Hoogerwerf (1991) enkele opmerkingen met betrekking tot evaluatie die vrijwel altijd opgeld doen. Zij wijzen drie, volgens hen, essentiële elementen uit hun definitie van evaluatie aan. Ten eerste betreft de evaluatie de *waarneming*

---

<sup>4</sup> Hoogerwerf kon in 1982 in een ‘paper’ vrijwel een compleet overzicht geven van het tot dan toe verrichte evaluatieonderzoek en de resultaten daarvan (Bressers, 2003)

van een situatie of verandering, en dus niet de daadwerkelijke situatie of verandering zelf. Degene die de evaluatie uitvoert neemt de situatie waar, dit is dus niet de situatie zelf. Evaluatie is dus nooit volledig objectief, maar wordt beïnvloed door de waarneming van de onderzoeker. Een tweede punt is het *vergelijken* van waarnemingen met criteria. Deze criteria kunnen velerlei van aard zijn, en zijn van grote invloed op de uitkomsten van de evaluatie. Of, zoals Bressers en Hoogerwerf (1991) het stellen: “Met een zeef scheidt men andere zaken uit de (beleids)poel dan met een emmer”. Het derde punt is het *verschil* tussen de situatie of de verandering als *onderwerp* van evaluatie. Doordat veranderingen kunnen worden geëvalueerd, is evaluatie een wezenlijk onderdeel geworden van de terugkoppelings- en bijsturingsmechanismen van het (overheids)beleid. Het evalueren van een bestaande situatie is veel moeilijker, er is vaak geen vergelijkingsmateriaal voorhanden waartegen de huidige situatie kan worden afgezet.

Patton (2001) geeft een belangrijke opmerking in het kader van dit onderzoek, namelijk met betrekking tot de functies van beleidsevaluatie. Patton (2001) ziet niet alleen het *bepalen* van de effectiviteit van beleid als een belangrijke functie, maar ook het *bijdragen* aan de effectiviteit van beleid. Bij de implementatie van nieuwe programma's en strategieën kan, doordat men bepaalde patronen herkent als gevolg van (eerder uitgevoerde) evaluaties, de effectiviteit van deze programma's en strategieën beter inschatten. Patton (2001) omschrijft dit als volgt: “*Being knowledgeable about patterns of program effectiveness allows evaluators to provide guidance about development of new initiatives, policies, and strategies for implementation. Such contributions constitute the conceptual use of evaluation findings.*” (p.333). Door gebruik te maken van de in dit onderzoek ontwikkelde methode voor het bepalen van de effectiviteit kan beleid wellicht effectiever worden opgesteld en kan de effectiviteit van nieuw of aangepast beleid worden verbeterd.

Op deze veronderstelling is echter ook kritiek, van onder andere Weiss (2002). Hij meent dat eerder uitgevoerde evaluaties geen garantie bieden voor de toekomst, omdat de omstandigheden continu aan veranderingen onderhevig zijn.

Bressers en Hoogerwerf (1984) zijn net als Patton van mening dat beleidsevaluatie een onmisbare functie heeft ten behoeve van terugkoppeling en bijsturing van beleid. Tegenwoordig wordt beleidsevaluatie dan ook als een belangrijk onderdeel gezien van het gehele beleidsproces. Het beleidsproces wordt door Hoogerwerf (1998) omschreven als “het dynamische verloop van handelingen en interacties met betrekking tot het beleid”.

## 2.5 Onderzoekstypen

Het onderzoekstype voor het evalueren van beleid hangt nauw samen met de eisen, effectiviteit, efficiëntie en legitimiteit, die gesteld worden aan beleid. Een onderzoeker wil vaak weten in welke mate aan (één of meer van) deze eisen wordt voldaan wil weten. Elke eis heeft min of meer zijn eigen onderzoekstype, hoewel er ook veel overlapping tussen de onderzoekstypen bestaat. Honigh (1985) geeft alleen een onderscheid aan tussen doelbereikingsonderzoek (goal attainment approach) en effectiviteitsonderzoek. Honigh merkt op dat zelfs deze twee in de literatuur nogal eens worden verward. Vooral het begrip effectiviteitsonderzoek wordt vaak onterecht gebruikt, wanneer een doelbereikingsonderzoek wordt bedoeld. Naast het onderscheid van Honigh zijn in de literatuur diverse onderzoekstypen terug te vinden die nog een stap verder gaan. De verwarring over het type onderzoek neemt hierbij navenant toe. In deze paragraaf worden de verschillende onderzoekstypen toegelicht om deze verwarring weg te nemen, althans hiertoe een poging te doen<sup>5</sup>.

---

<sup>5</sup> Voordat gekozen wordt voor een methode voor de beleidsevaluatie is het belangrijk om vast te stellen welk soort onderzoek uitgevoerd gaat worden. ‘Evaluators’ (degenen die de evaluatie uitvoeren) moeten er volgens Perrin (2001) naar streven om de simpelste methode te gebruiken, die mogelijk is om de evaluatie uit te voeren.

Van Hoesel (2003) geeft aan dat 'evaluatieonderzoek' op zich min of meer een overlappende term is voor de deelonderzoeken die daaronder vallen. Een van de onderzoekers die een helder overzicht geeft van onderzoekstypen is Bressers (2003). Allereerst onderscheidt hij het doelbereikingsonderzoek, het makkelijkste type onderzoek. Hierbij staat de vraag centraal in hoeverre bepaalde doelen zijn bereikt, zonder dat wordt nagegaan of de geconstateerde mate van doelbereiking ook werkelijk aan het gevoerde beleid is te danken. Men is snel geneigd geringschattend te spreken over de waarde van dit type onderzoek voor de evaluatie van beleid. Bressers (2003) is het hier slechts gedeeltelijk mee eens. Hij merkt terecht op dat de monitoring van "zowel de doelbereiking als de inzet van beleidsmiddelen" een grote stap vooruit is<sup>6</sup>. Niet alleen kan lering worden getrokken uit de resultaten, de voortkomende gegevens kunnen ook als basis dienen voor een effectiviteitsstudie.

Een benadering die hierbij aansluit is die van Arts (1998). Monitoring wordt ook door Arts als een type onderzoek gezien, met betrekking tot beleidsevaluatie, hoewel hij in minder harde termen spreekt van doelbereiking. Monitoring is in essentie dataverzameling, wat tot doel heeft om te bepalen of verwachte effecten zijn opgetreden en zo ja, in welke mate. Monitoring kan gezien worden als een activiteit om specifieke informatie over bepaalde variabelen (zowel ruimtelijk als socio-economisch) in de tijd en ruimte boven tafel te krijgen. Daarom heeft monitoring vaak betrekking op een reeks van waarnemingen en metingen van deze variabelen over een bepaalde tijdsperiode én met een bepaald doel. Monitoring zoals Arts dat voorstelt kan ook "doelstelling-vrij" zijn. Dit blijkt regelmatig voor te komen; in monitoringsrapporten worden ook ontwikkelingen en situaties genoemd die niet rechtstreeks verband houden met beleidsdoelstellingen. Arts ziet doelbereikingsonderzoek niet als volwaardig evaluatieonderzoek. Evaluatie begint volgens Arts pas bij een effectiviteitsonderzoek. Hoogerwerf (1998) merkt ook terecht op dat niet de doelbereiking maar juist de effectiviteit het echte beeld van de baten van het beleid oplevert.

Een stap verder in het beleidsonderzoek is het effectiviteitsonderzoek. Effectiviteitsonderzoek is onderzoek naar effecten. Hierbij wordt wel nagegaan wat de bijdrage van het beleid is geweest aan de doelbereiking. De causale relatie tussen het gevoerde beleid en de waargenomen situaties of ontwikkelingen staat hierbij centraal. Hierdoor wordt volgens Bressers (2003) het onderzoek veel ingewikkelder. Bressers (2003) merkt tevens terloops op dat niet alleen moet worden gemeten, maar ook dient te worden verklaard. Hoewel Bressers hier verder niet op ingaat, is dit wel van groot belang voor dit onderzoek, aangezien de kwaliteit van een effectiviteitsonderzoek wordt bepaald door de mate waarin wordt voldaan aan het verklaren van de effectiviteit!

Dit is, als onderdeel van het evaluatieonderzoek, geen sinecure. Het verklaren van succes of falen van beleid onderscheidt het evaluatieonderzoek van onderzoek naar effecten of monitoring. In de literatuur zijn auteurs te vinden die enkele handreikingen doen waarmee het succes van het beleid kan worden vastgesteld. Hoogerwerf (1998) meent dat voor het succes of falen van een overheidsbeleid verschillende factoren een rol spelen:

1. Doeleinden, waarbij onderscheid wordt gemaakt tussen doeleinden van beleidsuitvoerende instanties en die van de overige actoren, kunnen elkaar tegenwerken maar ook versterken;
2. Beschikbare capaciteit van middelen (zowel kwantitatief als kwalitatief) van de uitvoerende instantie;

---

De kans bestaat dat het hanteren van (te) moeilijken methoden resulteert in een wat Perrin (2001) 'methodological overkill' noemt. Hierdoor zullen de tijd en de kosten die met de evaluatie gepaard gaan toenemen, terwijl de weerstand en het cynisme tegenover evaluatie ook toenemen.

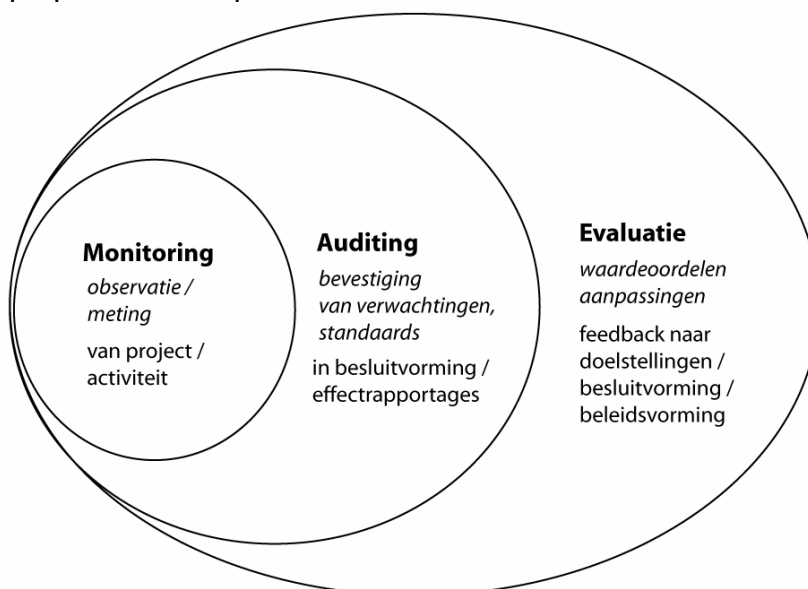
<sup>6</sup> "Steeds meer organisaties, niet alleen van overheid in engere zin, maar ook politie en universiteiten, vinden het normaal om belangrijke indicatoren van hun prestaties continu te monitoren" (Bressers 2003, p. 180)

3. De informatie van betrokken actoren schiet vaak tekort;
4. Machtsverdeling; de mate waarin deelnemers aan beleidsuitvoering in staat zijn hun eigen doeleinden na te streven.

Deze vier punten zouden bij een beleidsevaluatie moeten worden genoemd, bij het verklaren van het al dan niet optreden van effecten als gevolg van een bepaald beleid. Daarnaast speelt niet alleen de capaciteit van middelen een belangrijke rol, maar ook de capaciteit van de mensen. Bij het uitvoeren van beleid kan men er niet omheen dat menselijke interactie van grote invloed is.

Arts (1998) gebruikt een andere term om effectiviteitsonderzoek aan te geven, namelijk auditing. Auditing is een onafhankelijk, objectief onderzoek naar de uitkomsten van de uitvoering van het beleid in verhouding tot verwachtingen of standaards. Het belangrijkste verschil tussen auditing en monitoring is dat auditing betrekking heeft op het vaststellen en bevestigen van een voorspelling en/of vooraf vastgestelde criteria, standaards en procedures. Tussen de begrippen effectiviteitsonderzoek en auditing bestaat echter wel een belangrijk verschil. Een audit is volgens Arts, voor zover mogelijk, vrij van een waardeoordeel. Auditing kan derhalve gezien worden als een beperkte vorm van evaluatie, welke bedoeld is om bevestiging van vooraf gestelde verwachtingen te zoeken. In tegenstelling tot monitoring is auditing een eenmalige of periodieke gebeurtenis (geen continu proces).

Het volgende, meest complexe, onderzoekstype dat Bressers onderscheidt is het efficiëntieonderzoek. Hierbij staat de verhouding tussen de kosten en baten van een beleid centraal. Arts sluit hierbij aan met zijn omschrijving van evaluatie. Evaluatie (ex post) verwijst naar het proces van verzamelen, structureren, analyseren en beoordelen van data; het omvat zowel monitoring als auditing. Bij evaluatie worden expliciet subjectieve, beleidsgeoriënteerde oordelen gegeven over de effecten en effectiviteit van procedures en resultaten. Tevens dient hierbij een verklaring te worden gegeven over de waargenomen effecten en effectiviteit. Ex post evaluatie kan resulteren in extra metingen, aanpassingen of beslissingen met betrekking tot een project of beleidsplan.



**Figuur 2.3** Verhouding tussen monitoring, auditing en evaluatie

Arts geeft in Figuur 2.3 de verhouding tussen monitoring, auditing en evaluatie weer. Evaluatie omvat alles wat ook bij monitoring en auditing geldt en kan worden gezien als de ultieme vorm van evaluatieonderzoek. Auditing heeft ook betrekking op monitoring, maar wordt

gebaseerd op vooraf vastgelegde criteria of waarden. Monitoring houdt de observatie of meting van een project of activiteit in.

In wordt in Figuur 2.4 een overzicht gegeven van de eisen aan beleid in combinatie met de onderzoekstypen zoals die door Bressers en Arts worden gedefinieerd. Hierbij is gekeken naar de eisen die door beiden aan de onderzoekstypen worden gesteld.

Eisen aan beleid	Onderzoekstype Bressers	Onderzoekstype Arts
Doelbereiking	Doelbereikingsonderzoek	Monitoring
Doeltreffendheid	Effectiviteitsonderzoek	Auditing
Doelmatigheid	Efficiëntieonderzoek	Evaluatie
Legitimiteit	X	X

**Figuur 2.4** Overzicht eisen aan beleid en onderzoekstypen

In Figuur 2.4 worden de eisen aan beleid gekoppeld aan de typen onderzoek zoals die door Bressers (2003) en Arts (1998) worden onderscheiden. Hoewel doelbereiking niet als harde eis is opgenomen, wordt deze toch meegenomen omdat doelbereikingsonderzoek en/of monitoring onderdeel uitmaken van het beleidsonderzoek. In Figuur 2.4 wordt ook verduidelijkt dat evaluatieonderzoek zoals dat door Arts en Bressers wordt gezien min of meer uit een soort beslisboom bestaat. Daar wordt in de literatuur weinig op gewezen. Efficiëntieonderzoek (Bressers) of evaluatieonderzoek (Arts) is pas mogelijk wanneer ook inzicht is in doelbereiking en effectiviteit. Effectiviteit is pas te bepalen wanneer ook inzicht is verkregen in de doelbereiking. Om doelbereiking te kunnen bepalen moet eerst worden vastgesteld wat de doeleinden van het desbetreffende beleid waren. Deze opvatting wordt in de methode van onderzoek verder uitgewerkt.

Wat verder opvalt is dat legitimiteit meer als een onderdeel van het evaluatieonderzoek wordt meegenomen. Er bestaat geen type onderzoek dat alleen gericht is op de legitimiteit van beleid. Legitimiteit kan dan ook meer gezien worden als factor die van invloed is op de doeltreffendheid en vooral doelmatigheid van beleid dan een op zichzelf eis. Beleid moet uiteraard zo breed mogelijk gedragen worden, maar wanneer dit niet gebeurt dan heeft dat ook een negatieve invloed op de doelmatigheid van beleid. Als een evaluatieonderzoek wordt gedaan naar doelmatigheid van het desbetreffende beleid, dan zou uiteindelijk als oorzaak ook gebrek aan legitimiteit naar moeten komen. De vraag is dus of legitimiteit wel als harde eis aan beleid moet worden gesteld, aangezien dit min of meer onderdeel uitmaakt van doelmatigheid.

Wat hierbij nog relevant is wat door Dekker (2004) wordt aangegeven: de wisselwerking tussen de kwaliteit van beleid en de evaluatie. Hoe beter het beleid, des te beter vaak ook de evaluatie uitgevoerd kan worden. Tevens geeft hij aan dat wanneer bij het opstellen van beleid stilgestaan wordt bij de evalueerbaarheid ervan, dit voor zowel het beleid als de evaluatie voordeel biedt. Het functioneren van het beleid kan beter worden beoordeeld, terwijl ook de evaluatie beter uitgevoerd kan worden. Tot slot geldt ook dat wanneer de evaluatie duidelijk aangeeft op welke punten beleid tekortschiet of voldoet, men daar het beleid mee kan verbeteren.



### 3 Uitvoering van evaluatieonderzoek

#### 3.1 Inleiding

Iedereen evalueert, velen evalueren overheidsbeleid en slechts weinigen evalueren overheidsbeleid wetenschappelijk (Bressers, 2003). Volgens Hoogerwerf en Zoutendijk (1991) kan daar nog aan worden toegevoegd dat zeer weinig mensen overheidsbeleid evalueren op een manier die geheel voldoet aan kwaliteitseisen vanuit de wetenschap én vanuit de beleidspraktijk.

Het uitvoeren van een evaluatieonderzoek vereist het doorlopen van een aantal stappen om tot een gedegen oordeel te kunnen komen. Deze fasen worden in dit hoofdstuk beschreven. Op het eerste gezicht lijkt dit wellicht overbodig, aangezien dit onderzoek niet zelf een beleidsevaluatieonderzoek betreft, maar om tot een goed oordeel te kunnen komen over de beschouwde beleidsonderzoeken is het noodzakelijk om inzicht te hebben in de methoden van beleidsonderzoek. Hierop wordt immers het oordeel gebaseerd.

Volgens Bressers (2003) kunnen drie fasen worden onderverdeeld, namelijk de voorbereidingsfase, de uitvoeringsfase en de afrondingsfase. Deze fasen worden hier achtereenvolgens toegelicht, aan de hand van Bressers' onderscheid en in de literatuur verzamelde inzichten.

#### 3.2 Voorbereidingsfase

In de voorbereidingsfase wordt het onderwerp van de evaluatie gekozen evenals de criteria waarop gaat worden beoordeeld. Beleidsevaluatie heeft betrekking op het beoordelen van de waargenomen inhoud, processen of effecten van een bepaald beleid aan de hand van bepaalde criteria (Bressers en Hoogerwerf, 1984). Deze criteria waarop beleid wordt beoordeeld kunnen, wanneer beleid als een doelbewust streven wordt beschouwd, in belangrijke mate worden afgeleid van de doelstellingen van het betreffende beleid<sup>7</sup>.

De selectie van criteria lijkt eenvoudig, maar zowel Bressers (1984) als Lulofs en Schuddeboom (1991) zijn van mening dat er in de praktijk complicerende factoren spelen bij de criteriakeuze. De meest vanzelfsprekende, rechttoe-rechtaan toepassing van evaluatieonderzoek is mogelijk wanneer de betrokkenen (beleidsvoerders en onderzoekers) zich vooraf vastleggen op evaluatiecriteria en –standaarden. Evaluatiecriteria zijn dan de dimensies, geoperationaliseerd tot variabelen, waarop geëvalueerd wordt, standaarden zijn bepaalde waarden – b.v. grensscores – gedefinieerd op de criteria. Bij toepassing van de evaluatiestandaarden is er sprake van een succes als de uitkomst op het criterium boven de standaard ligt een mislukking als beneden de standaard gescoord wordt (of vice versa).

Het ligt voor de hand dat de onderzoeker die de evaluatie uitvoert en de mate van doelbereiking wil vaststellen, de beleidsdoelen als bron van evaluatiecriteria kiest. Beleid, als zijnde een min of meer samenhangend geheel van doeleinden en middelen, bevat zelf ook doeleinden. De indicatoren van doelbereiking worden afgeleid van de doeleinden die binnen het kader van het officiële overheidsbeleid expliciet zijn geformuleerd (Bressers, 1984).

Zelfs wanneer de keuze voor de criteria is gemaakt, zijn nog niet alle problemen opgelost. Juist het uitgaan van de officiële beleidsdoelen bij de keuze van evaluatiecriteria heeft in de literatuur veel kritiek gekregen. Deze kritiek is dat de officiële beleidsdoeleinden niet overeen komen met de werkelijke doeleinden van de beleidsvoerders (Bressers, 1984). Dit heeft

---

<sup>7</sup> Hierbij schuilt volgens Bressers en Hoogerwerf het gevaar dat beleidsvoerders naast hun officiële (formele) doeleinden ook informele doeleinden hebben die zij nastreven, waardoor zij beleid positief blijven beoordelen, terwijl de beleidsevaluatie op basis van officiële doeleinden negatief uitpakt.



volgens Lulofs en Schuddeboom (1991) echter niet te betekenen dat het gebruik van officiële beleidsdoeleinden als evaluatiecriteria verworpen moet worden. Wanneer beleidsvoerders andere doelstellingen nastreven, dan kunnen de officiële doeleinden een belangrijke mate de legitimatie van het beleid in een democratisch proces vormen, waardoor ze juist uitstekend als maatstaf kunnen worden gebruikt (Hoogerwerf, 1998)

Een ander probleem heeft te maken met de wijze waarop de officiële doeleinden zijn geformuleerd. De doeleinden kunnen zo vaag zijn geformuleerd dat zij moeilijk zijn te indiceren en daardoor de doelbereiking moeilijk is te meten of de doeleinden zijn geformuleerd in termen van elementen die slechts zeer indirect door het beleid worden beïnvloed. Daarnaast kunnen de doeleinden onrealistisch hoog of laag gesteld zijn, of zijn de doeleinden veranderlijk of tegenstrijdig (Bressers, 1984).

Het is eigenlijk opvallend dat er in de meer onderzoekstechnische literatuur over evaluatieonderzoek zelden rekening gehouden wordt met de mogelijkheden dat er vooraf criteria en standaarden zijn vastgelegd. Volgens Scheerens (1983) dient, wanneer evaluatieonderzoek strikt serieus wordt genomen, ook door de beleidsbeslissers, het vooraf specificeren van criteria en standaarden zelfs als een noodzakelijke voorwaarde te worden beschouwd. In de praktijk zal deze eis vaak te sterk worden gevonden. Lulofs en Schuddeboom (1984) zien dit niet als harde eis, maar doen wel de aanbeveling “om in overleg met de opdrachtgever de evaluatiecriteria (beleidsdoelen) en hun operationalisaties vast te leggen” (p.77).

Tot slot nog dit: hoe ruimer het onderwerp wordt gekozen, hoe complexer het onderzoek zal zijn. In de voorbereidingsfase dient daarom tevens de informatiebehoefte te worden vastgesteld en een methode van onderzoek te worden gekozen. Het komen tot een heldere, volledige en juiste probleemformulering en – daaraan gekoppeld – een goede probleemstelling en vraagstelling voor het onderzoek is één van de meest cruciale fasen in een evaluatieonderzoek ex post: het hele onderzoek wordt er in feite mee gestuurd.

### 3.3 Uitvoeringsfase

De uitvoeringsfase is volgens Bressers (2003) de belangrijkste fase in evaluatieonderzoek. Deze fase bestaat uit vier stappen:

#### **1. Beschrijven van de beleidsinhoud aan de hand van een “impact model” of beleidstheorie.**

Om een beleid goed te kunnen evalueren, moet men allereerst een duidelijk beeld verwerven van wat het beleid precies inhoudt. Verschillende auteurs wijzen op de noodzaak om bij het effectiviteitsonderzoek gebruik te maken van een zogenaamd “impact model”. In een dergelijk impact model worden causale relaties tussen de toepassing van de middelen van een beleid en het bereiken van de beleidsdoelen geëxpliciteerd. Bij onderzoek naar de effectiviteit van een regeling staat de relatie tussen een instrument (de instrumentvariabele) en het beoogde effect (de doelvariabele) centraal. Echter, op deze relatie werken veelal andere factoren in (zie Figuur 2.1). Een onderzoeker inventariseert daarom alle relevante factoren en relateert ze vervolgens aan elkaar in een analyseschema. Een manier om relevante factoren te achterhalen is om de door de beleidsvoerders gehanteerde beleidstheorie te reconstrueren<sup>8</sup>. Een beleidstheorie is een theorie die de relaties tussen het toepassen van beleidsmiddelen (de gekozen oorzaken, de onafhankelijke variabelen) en de gewenste beleidseffecten (de beoogde gevolgen, de onafhankelijke variabelen) aangeeft (Bressers, 1984). Deze beleidstheorie kan als ‘input’ worden gezien voor het “impact model”. Bressers (1984) meent dat een overzicht van het beleidsveld essentieel is, omdat bij effectiviteitsonderzoek juist deze relaties worden onderzocht. Alleen wanneer men de inhoud van het beleid en het betreffende beleidsveld als een samenhangend geheel beschrijft, is het mogelijk om een overzicht krijgen

<sup>8</sup> Het beeld van beleidsbepalers strookt vaak niet met de werkelijkheid (Bressers, 2003).

van de relaties, die voor het effectiviteitsonderzoek van belang zijn. Daarom zou bij elk evaluatieonderzoek allereerst het beleidsveld moeten worden uitgewerkt.

## **2. Vaststellen van de doelbereiking, oftewel de mate waarin aan de gestelde criteria wordt voldaan**

Hierbij gaat het om twee substappen, namelijk (1) het operationaliseren (meetbaar maken) van de criteria<sup>9</sup> en (2) de dataverzameling<sup>10</sup>. Het operationaliseren van doeleinden, vraagt in de praktijk vaak om een vertaling van de voor het evaluatieonderzoek geëxpliciteerde doeleinden in meetbare termen. Bij beleid dat gericht is op een maatschappelijk probleem moeten de veelal abstracte en vaag geformuleerde wensen die bij actoren leven over een maatschappelijk probleem worden geoperationaliseerd in criteria en indicatoren (Ministerie van Financiën, 1993). Er moeten indicatoren worden gekozen waarmee de criteria in waarneembare grootheden worden uitgedrukt.

Indien een doelstelling bijvoorbeeld luidt 'sterke verbetering van de leefbaarheid', dan moet eerst duidelijke worden gemaakt wat 'leefbaarheid' inhoudt. Vervolgens moet de 'sterke verbetering' door de onderzoeker worden indiceren in meetbare termen. Aan deze operationalisering wordt bij evaluatieonderzoek in de praktijk lang niet altijd voldaan, terwijl dat volgens Lulofs en Schuddeboom (1991) een eis moet zijn wanneer men wil komen tot een goed evaluatieonderzoek. Hoogerwerf (1998) daarentegen is het hier slechts gedeeltelijk mee eens. Wat er volgens hem uit hoofde van wetenschappelijke degelijkheid echt toe doet, is dat men zijn criteria expliciet maakt, zodat de lezer van het onderzoeksverslag weet waarop het oordeel is gebaseerd. In dit onderzoek wordt meer de filosofie van Lulofs en Schuddeboom gevolgd, aangezien hierbij daadwerkelijk uitspraken kunnen worden gedaan over het bereiken van de doelstellingen.

Bij het operationaliseren van doeleinden zijn tenminste drie onderdelen van belang, namelijk de relatie tussen doel en indicator, de dimensies van de gekozen indicator en interveniërende factoren. Deze drie onderdelen moeten worden aangegeven in het beleidsevaluatieonderzoek. Het genoemde "impact model" is hiervoor een goed uitgangspunt.

Bij het operationaliseren van doelstellingen doet zich het volgende gevaar voor. Als vooraf duidelijke doelstellingen worden aangegeven, dan kan evaluatie bestaan uit nagaan of de doelen worden bereikt. In een dergelijke benadering ligt het accent vaak op het meetbaar maken daarvan. Dat brengt verschillende bezwaren met zich mee: van de intenties van programmamakers blijft bij het meetbaar maken soms weinig over (juist de belangrijkste doelen blijken moeilijk meetbaar), tijdens de uitvoering van het programma kunnen allerlei neveneffecten optreden waarmee van tevoren geen rekening is gehouden, en de uitvoerende instanties worden als het ware vastgeprikt op bepaalde doelstellingen: dat vergroot de mogelijkheden voor controlerende instanties om greep te hebben op wat er zich afspeelt, en verkleint de ruimte voor vernieuwingen. Het probleem van neveneffecten bracht Scriven (1974) ertoe om de gedachte van 'goal-free evaluation' te lanceren. In zijn suggestie zat zelfs opgesloten dat de evaluatieonderzoeker totaal onbekend blijft met de officiële doelen in kwestie. Dat zou hem onbevooroordeeld houden. Probleem is echter wel dat toch gekozen moet worden waar wel en niet op te letten in onderzoek (Stokking, 1991). Om een onderzoeker helemaal onvoorbereid het onderzoek in te laten gaan, lijkt niet optimaal. De onderzoeker zal daardoor niet weten welke informatie hij nodig heeft en daardoor kan tijdens het verloop van het onderzoek blijken dat het doel dat hij nastreeft, het inzichtelijk maken van de doelbereiking, effectiviteit en efficiëntie van een beleid, niet haalbaar is omdat relevante informatie ontbreekt.

<sup>9</sup> Dit dient zeer zorgvuldig te gebeuren, aangezien het kan leiden tot een beperking of verschuiving van het criterium zoals het oorspronkelijk was geformuleerd.

<sup>10</sup> Men moet kritisch zijn ten opzichte van de gebruikte gegevens, de kwaliteit van gegevens moet voorop staan.

### 3. Vaststellen van de doeltreffendheid, oftewel de effectiviteit van beleid vaststellen (zie Figuur 2.2)

Voor het vaststellen van de doeltreffendheid is het noodzakelijk om gegevens te verzamelen over de criteria die in de voorgaande stappen zijn gekozen en geoperationaliseerd. Hiervoor zijn zoveel verschillende methoden in omloop, dat het onmogelijk is deze allemaal hier te behandelen. Bressers (2003) meent dat er “een gereedschapskist van methoden en technieken” nodig is om te kunnen bepalen in hoeverre het beleid daadwerkelijk heeft bijgedragen aan de mate van doelbereiking. Wat van belang is bij deze methoden en technieken is data die verzameld dienen te worden.

Het verzamelen van de data is niet zo eenvoudig als het in eerste instantie lijkt. Bressers (1984) komt tot de conclusie dat “de onderzoekers meer dan eens ondervonden dat hun aanvankelijk overwegend kwantitatieve en (quasi-) experimentele methoden eisen stelden aan het onderwerp van het onderzoek (de onderzoekseenheden en de beschikbare gegevens), waaraan niet kon worden voldaan”. Hierdoor ontstond steeds meer de behoefte aan andere, vaak kwalitatieve, methoden. Om problemen te kunnen opvangen die ontstaan wanneer de omstandigheden een volledige experimentele onderzoeksopzet onmogelijk maken, zijn verschillende alternatieven, quasi-experimentele onderzoeksontwerpen, ontwikkeld. Deze zijn allemaal gebaseerd op het maximaal benutten van informatie uit die vergelijkingsbases die bij het onderzoeksontwerp goed zijn aan te brengen. De experimentele methode is er op gericht de mate van het effect van een bepaalde maatregel vast te stellen. Dit gebeurt door de vergelijking van een experimentele groep, die onderworpen wordt aan de maatregel, met een controlegroep waarvoor dat niet geldt (Ziegelaar, 2004). Bressers (1984) stelt dat één van de bekendste van deze quasi-experimentele onderzoeksontwerpen het longitudinale of tijdsreeksen onderzoek is. Hierbij wordt het verloop van een afhankelijke variabele nagegaan over een bepaald tijdsbestek. Daarbij wordt bezien in hoeverre de prikkel van het beleid samenhangt met een wijziging in dit verloop en in hoeverre deze wijziging als een invloed van het beleid kan worden geïnterpreteerd. Probleem hierbij is dat over de gehele periode gedurende het tijdsbestek voldoende gegevens beschikbaar moeten zijn. Er mogen dus geen data ontbreken of randvoorwaarden veranderen tijdens het tijdsbestek. Dit blijkt in de praktijk echter vaak wel het geval, waardoor dit type onderzoek, wetenschappelijk gezien, vrijwel niet is uit te voeren.

Patton (2001) en Millett bekeken de ‘beste werkwijzen’ en ‘geleerde lessen’ met betrekking tot de dataverzameling zoals die in veel evaluatieonderzoeken zijn terug te vinden en onder een grote noemer worden geschaard. Hun conclusie is dat de items die onder dezelfde noemer werden gezet, zoveel breder en zoveel meer omvattend waren, dat de conclusies vaak vrij waren van enige consistentie. Dé beste werkwijze of dé geleerde les met betrekking tot de dataverzameling bestaat volgens hen dus niet. Perrin (2001) drukt dit als volgt uit: *“scientifically ‘perfect’ designs are almost never possible [...] A common mistake is to try to impose an idealized design upon a situation, rather than starting by prioritizing critical information needs and only then identifying the simplest and most appropriate means of addressing them.”* (p.253)

Volgens Perrin (2001) is het dus belangrijker om in het beginstadium vast te stellen welke essentiële informatie of data boven tafel moet worden gehaald<sup>11</sup>, dan het aangeven welke methode zal worden gebruikt. Lulofs en Schuddeboom (1991) zijn het hier mee eens, al omschrijven zij het anders: “wanneer wordt overgaan tot de daadwerkelijke dataverzameling, dan is van belang om in het oog te houden uit welke bronnen kan worden geput, welke

<sup>11</sup> Dit is onmogelijk wanneer de methodiek gevolgd wordt zoals die door Scriven (1974) wordt aangegeven: de onderzoeker het ‘goal-free-’onderzoek laten ingaan zonder dat hij weet wat er speelt. Hierdoor kan de onderzoeker immers nooit weten welk informatie hij nodig heeft.

tijdstippen relevant zijn voor de dataverzameling en welke technieken gebruikt moeten worden voor de dataverzameling.” Perrin (2001) voegt hier nog aan toe dat ‘de mensen’ en communicatieve vaardigheden minstens zo belangrijk zijn bij evaluatieonderzoek als vaardigheden in de onderzoekstechnieken.

Dit betekent dat niet zozeer het onderzoeksontwerp en de keuze van de methode de belangrijkste onderdelen zijn bij evaluatieonderzoek, maar de verzamelde data en de onderzoeker die het evaluatieonderzoek uitvoert. Het maakt weinig uit op welke manier de data wordt verzameld, als de data maar voldoet aan enkele eisen. In de praktijk blijkt dit vaak over het hoofd te worden gezien. Te veel tijd wordt besteed aan de methode van onderzoek, terwijl onduidelijk is welke data men eigenlijk wil verzamelen, welke informatie beschikbaar is en hoe men hier mee om zal gaan.

Lulofs en Schuddeboom (1991) geven een voor dit onderzoek enkele belangrijk punten aan, die zich doen gelden bij de dataverzameling. Zij zijn van mening dat wanneer met bestaande gegevens wordt gewerkt, het van belang is om na te gaan hoe deze gegevens werden verkregen. Dit om te voorkomen dat er een grotere of geringere exactheid aan wordt toegeschreven dan ze verdienen en om de kans op een onjuiste interpretatie te verminderen. Daarnaast is het wel belangrijk om, ondanks dat gebleken is dat er geen universeel geldende regels voor de correcte wijze en tijdstippen van de dataverzameling ten behoeve van evaluatieonderzoek zijn, algemene eisen van uitvoerbaarheid, accuraatheid en bruikbaarheid van het onderzoek in het oog worden gehouden. Lulofs en Schuddeboom (1991) hebben enigszins algemeen geldende ‘regels’ van Patton voor het verzamelen van data uitgewerkt in vijf vuistregels.

Door het Ministerie van Financiën (1993) wordt aangegeven dat de keuze voor een bepaalde dataverzamelingstechniek in het algemeen ingegeven zal zijn door enerzijds de mogelijkheden en beperkingen van de onderzoeker en anderzijds door de mogelijkheden en beperkingen van de techniek. Volgens Arentsen (1984) is het nodig in het effectiviteitsonderzoek om op ten minste twee tijdstippen data te verzamelen: voordat de beleidsprikkel is toegediend en na de toediening van de beleidsprikkel. Dit blijkt in de praktijk vaak moeilijk realiseerbaar (Dekker, 2004): het beleid wordt geëvalueerd op het moment dat het al in uitvoering is. Hierdoor vervalt de mogelijkheid om data te verzamelen voor het moment dat de beleidsprikkel is toegediend. Of de groep die tijdens het onderzoek met het beleid in aanraking is gekomen, blijkt niet representatief voor de gehele doelgroep. Tegenwoordig, nu het belang van beleidsevaluatie meer is doorgedrongen, wordt hier in enkele gevallen rekening mee gehouden door in het eerste stadium een zogenaamde nulmeting uit te voeren.

#### **4. Verklaren van de effectiviteit van het beleid, oftewel het bepalen van de doelmatigheid, met als doel de bijdrage van het beleid aan de doeltreffendheid te vergroten (kosten-baten)**

Wanneer is vastgesteld welke effecten zijn opgetreden en in hoeverre dit te danken is het uitgevoerde beleid, dan is de onderzoeker bij evaluatieonderzoek nog niet klaar. Hij zal ook een verklaring moeten geven over het succes of falen van het beleid. Onder andere Bressers (1984) geeft aan dat naast de vraag in hoeverre een beleid effectief is, ook de vraag moet worden beantwoord hoe het komt dat het beleid meer of minder effectief is. Deze informatie is nodig om te weten hoe men de beleidseffectiviteit zou kunnen verbeteren.

Bijna tien jaar later, in 2003, komt Bressers met de opmerking dat juist hier het evaluatieonderzoek van de afgelopen jaren tekortschiet: “deze verklaringen zijn op zichzelf plausibel en gelden voor het specifieke beleid, maar leren weinig over hoe te handelen op een andere plaats, bij een iets ander beleid, of op een ander moment”. Hij meent zelfs dat op dit moment enige wetenschappelijke uitspraken over de doelmatigheid van beleid onmogelijk zijn, aangezien de theoretische kennis over de werking van beleid onvoldoende is!

Deze laatste opmerking is voor dit onderzoek relevant, aangezien er gezocht gaat worden naar een evaluatiemethode die ook geschikt is om de beleidsvorming in een vroeg stadium te kunnen verbeteren. Wanneer de redenering van Bressers wordt gevolgd, dan is de verwachting dat dit niet mogelijk is, tenminste niet volgens de wetenschappelijke eisen.

### 3.4 Afrondingsfase

De laatste fase in het evaluatieonderzoek die Bressers onderscheidt is de afrondingsfase. Hierbij staat de rapportage van de resultaten centraal, met als doel de terugkoppeling van evaluatiebevindingen naar beleid te verbeteren, oftewel de bijsturing te bevorderen. Om deze terugkoppeling te kunnen uitvoeren is het noodzakelijk om conclusies te verbinden aan het evaluatieonderzoek.

Het trekken van de conclusie(s), blijkt in de praktijk een complexe aangelegenheid. Allereerst blijkt beantwoording van de vraag naar de effectiviteit van het beleid meestal niet eenvoudig. Bressers (1984) meent zelfs dat de vraag naar potentiële effecten van een beleid nooit in zijn geheel zal kunnen worden beantwoord. Van elk beleid zijn namelijk ontelbare van zulke effecten denkbaar. Voor dit onderzoek is deze opmerking van belang. Bressers geeft dus aan dat een effectiviteitsonderzoek nooit voor honderd procent de effecten van een beleid aan kan geven. Dit betekent dat in de praktijk zal blijken dat in een effectiviteitsonderzoek nooit alle effecten van een beleid zijn opgenomen. Enige terughoudendheid ten aanzien van de uitkomsten van effectiviteitsonderzoek is dan ook geboden wanneer het gaat om het verbinden van harde conclusies en het nemen van maatregelen op basis van het effectiviteitsonderzoek.

Door het Ministerie van Financiën (1993) wordt dit probleem ook onderkend. Daadwerkelijke beleidseffecten zijn vaak moeilijk te herkennen. Hierdoor bestaat de kans dat de onderzoeker in de verleiding komt om conclusies te verbinden met betrekking tot (de uitvoering van) beleid, door effecten mee te nemen die er eigenlijk niet toe doen, of juist effecten niet mee te nemen als deze lastig zijn te herkennen, of überhaupt niet worden herkend. Dit is het probleem van de identificatie van effecten. Onderzoek naar effecten van beleid houdt daarom altijd een keuze in van de mogelijke effecten waarop men let, of de keuze nu openlijk wordt gemaakt of niet.

Het omzetten van de verzamelde informatie in 'harde' conclusies blijkt bij evaluatieonderzoek vaak moeilijk te zijn. Dit kan verschillende oorzaken hebben, zoals het ontbreken van bepaalde informatie, of een aanpassing van het beleid in de onderzoeksperiode (Bressers, 1984).

Een probleem is het verschil in opvatting over de conclusie: zoals eerder gesteld wordt bij evaluatieonderzoek dikwijls de conclusie getrokken dat een bepaald effect is opgetreden, terwijl het 'hoe en waarom' verder niet is uitgezocht, terwijl dit toch het een van de belangrijkste elementen is bij evaluatieonderzoek (het vaststellen van de bijdrage van een beleid bij het al dan niet optreden van effecten). Consequentie hiervan is dat de kans aanwezig is dat door de opdrachtgever en andere betrokkenen onterecht maatregelen worden genomen op basis van de conclusies zoals die door de onderzoeker in het evaluatieonderzoek zijn opgenomen.

Een interessant aspect bij het trekken van de conclusies wordt aangegeven door het Ministerie van Financiën. Het trekken van de conclusies dient volgens het Ministerie van Financiën (1993) beperkt te blijven tot feitelijke ondervindingen, en moet niet in gaan op de discussie over de vraag of doeleinden van het beleid het nastreven waard zijn en of zij afleidbaar zijn uit abstracte waarden als rechtvaardigheid, gelijkheid, vrijheid, democratie of solidariteit. Dit behoort tot het terrein van de politiek. Het zijn bij uitstek normatieve vragen die niet empirisch onderzoekbaar zijn. Een onderzoeker kan en mag zich daar niet mee bemoeien volgens het

Ministerie van Financiën (1993). De onderzoeker moet zich dus beperken tot wat hij tijdens het onderzoek heeft waargenomen.

### 3.5 Valkuilen en problemen bij evaluatieonderzoek

Zoals al is gebleken spelen bij evaluatieonderzoek in de praktijk verschillende problemen. Naast de genoemde problemen om effecten te identificeren spelen ook nog andere problemen een belangrijke rol. In de literatuur zijn hier diverse voorbeelden van te vinden. In deze paragraaf worden de meest voorkomende kort beschreven. Deze problemen en valkuilen zijn meegenomen in de methodiek die is opgesteld voor het bepalen van de kwaliteit van evaluatieonderzoek.

Allereerst kan algemeen gesteld worden dat de wisselwerking tussen beleid en onderzoek geoptimaliseerd dient te worden om een goed evaluatieonderzoek uit te voeren. Stufflebeam (1971) meent dat problemen bij evaluatieonderzoek meestal voortkomen uit gebrekkig inzicht in besluitvormingsprocessen en de relatie tussen evaluatieonderzoek en besluitvorming.

Daarnaast spelen problemen bij evaluatieonderzoek die verband houden met de opzet van het onderzoek. De veel gebruikte (**quasi-)experimentele** opzet is aan steeds meer kritiek onderhevig (Ziegelaar, 2004). Bij een (quasi-)experimentele opzet wordt het verloop van een afhankelijke variabele nagegaan over een bepaald tijdsbestek. Daarbij wordt bezien in hoeverre de prikkel van het beleid samenhangt met een wijziging in dit verloop en in hoeverre deze wijziging als een invloed van het beleid kan worden geïnterpreteerd. De kritiek spitst zich met name toe op de problemen die zich voordoen met betrekking tot de data, zoals gebrekkige data, niet beschikbare data, uitval uit de steekproef en onvergelijkbaarheid van controle- en experimentele groep (Ziegelaar, 2004).

Volgens Scheerens (1983) zijn de bekendste problemen bij evaluatieonderzoek de klachten over de **onduidelijke doelstellingen** van de te evalueren beleidsprogramma's. Volgens het Ministerie van Financiën (1993) kunnen deze problemen die zich toespitsen op de doelstellingen in een beleid als volgt worden samengevat:

1. Doelstellingen van beleid zijn vaak vaag en niet gekwantificeerd.
2. Doelstellingen van een beleid zijn soms symbolisch. Symbolische doelstellingen zijn niet primair gericht op feitelijke veranderingen. Ze vervullen vooral een retorische functie.
3. De verschillende belanghebbenden die bij een bepaald beleid zijn betrokken, kunnen er verschillende doelstellingen op na houden. Soms zullen die verenigbaar zijn, soms niet. Het kan gebeuren dat de verschillende belanghebbenden elkaar vinden in vage of symbolische doelstellingen, waaraan "niemand zich een buil kan vallen".
4. De doelstellingen van een beleid kunnen na invoering van het beleid door de tijd heen veranderen. Het onderscheid tussen enerzijds oorspronkelijke en huidige doelstellingen en anderzijds formele en materiële doelstellingen leidt tot vier combinaties. Bij een evaluatie moet voor onderzoekers en beleidsvoerders duidelijk zijn over welke doelstellingen men het heeft.
5. Het niveau van de doelstellingen kan worden gemanipuleerd. Een hoge doelstelling is moeilijk te bereiken, maar is aantrekkelijk om het beleid geaccepteerd en vastgesteld te krijgen. Bij de evaluatie zal echter vaak blijken dat de doelstelling onrealistisch is, zoals ook Bressers (1984) concludeert.

Indien men onderzoek wil doen naar de effectiviteit van beleid moet men voor de hiervoor genoemde problemen een oplossing zien te vinden. Beslissend daarbij blijft het zo nauwkeurig mogelijk specificeren en vaststellen van de doelstellingen.

Deze problemen met betrekking tot doelstellingen spelen volgens Scheerens (1983) bij overheidsbeleid nog sterker, aangezien het politieke karakter van de beleidsvoering ervoor zorgt dat de beleidsvoerders (politici) zich niet willen vastleggen op operationele

doelstellingen. Dit kan volgens Scheerens verschillende oorzaken hebben: (1) door doelstellingen vaag en algemeen te houden zou de mogelijkheid om compromissen te sluiten groter zijn, en (2) men beveiligt zich op deze manier tegen eventuele onwelgevallige uitkomsten van het evaluatieonderzoek, namelijk door achteraf te kunnen stellen dat de onderzoeksuitkomsten geen betrekking hebben op wat er bedoeld werd. Ook kan het politieke klimaat een rol spelen bij de zinvolle toepassing van evaluatieonderzoek.

Een ander probleem dat door Scheerens (1983) wordt aangegeven is dat het lastig kan zijn om **beleidsvoerders te identificeren**. Daarnaast speelt ook nog het feit dat er altijd andere informatie beschikbaar is naast de uitkomsten van beleidsgericht onderzoek, die, ondanks dat deze informatie niet altijd op waarheid hoeft te berusten, de discussie omtrent het onderzochte beleid kan beïnvloeden.

**Indicatoren** die worden gebruikt in een evaluatieonderzoek kunnen ook voor problemen zorgen. Perrin (2002) wijt dit aan het misbruik van indicatoren. Hij geeft een achttal punten aan, waarin het ineffectieve en misbruik van indicatoren voor het meten van de 'performance' van beleid ligt verscholen: (1) verschillende interpretaties van dezelfde voorwaarden en/of concepten, (2) misplaatste doelen, (3) gebruik van betekenisloze en irrelevante meetinstrumenten, (4) verwarring tussen kostenbesparing en kostenverschuiving, (5) onduidelijke verschillen tussen subgroepen door verkeerde integratie van indicatoren, (6) beperkingen van objectgebaseerde benaderingen van evaluaties, (7) onbruikbaarheid van indicatoren voor besluitvorming en verdeling van middelen, en (8) de inconsistentie tussen beperkte focus op 'meten' en het grotere publieke inzicht van decentralisatie en verdeling van autoriteit.

De Vries (2000) komt met een ander interessant probleem, namelijk het zogenaamde probleem van de **dubbele hermeneutiek**. Dit houdt in dat niet alleen de onderzochte sociale werkelijkheid een optelsom is van subjectieve werkelijkheidsdefinities, maar dat ook de waarneming van de onderzoeker een gekleurde is. "De onderzoeker, met andere woorden, interpreteert interpretaties" (p.21). Wanneer de onderzoeker zich hiervan bewust is, dan is dit volgens de Vries geen probleem, maar vaak blijkt dit niet het geval te zijn, waardoor te veel waarde aan de uitkomsten van metingen of waarnemingen wordt gehecht.

Daarnaast speelt het probleem van onduidelijkheid omtrent de **uitvoering** van evaluatieonderzoek. Zonder dat alle fasen van evaluatieonderzoek officieel worden doorlopen wordt toch vaak beleidsevaluatie uitgevoerd, door bijvoorbeeld regeringspartijen, de pers en doelgroepen van beleid (burgers). Deze 'gewone' beleidsevaluaties zijn volgens Hoogerwerf (1998) vaak niet toereikend. Volgens hem doordat praktijkkennis alleen niet voldoende is en wel om de volgende redenen:

- De meeste insiders zijn (vaak) ook belanghebbenden. Een subjectief beeld van de werkelijkheid is het gevolg;
- De kennis aan de basis dringt lang niet altijd door tot aan de ambtelijke of politieke top;
- De kennis van actoren uit de beleidspraktijk vertoont belangrijke hiaten. Het is vaak onduidelijk in hoeverre beleid heeft bijgedragen aan de doelstelling, met andere woorden wat de effectiviteit van het beleid daadwerkelijk is;
- De actoren uit de beleidspraktijk beschikken vaak niet over de mogelijkheid systematisch te vergelijken.

Dit punt van kritiek is van belang bij het trekken van conclusies aan de hand van een beleidsevaluatieonderzoek. Duidelijk moet zijn wie opdrachtgever en –nemer zijn, zodat men kan bepalen of de resultaten van de evaluatie wel of niet voldoende onderbouwd zijn.

Door Hoogerwerf en Zoutendijk (1991) zijn de meest genoemde problemen rond de beleidsevaluatie een tekort aan evaluatie, onvoldoende kwaliteit van de evaluatie, te weinig gebruik van evaluatierapporten bij de beleidsvorming en onvoldoende duidelijkheid van de

beleidsdoeleinden waaraan evaluatiecriteria kunnen worden ontleend. Maar waar liggen de oorzaken voor het gebrek aan kwaliteit van de evaluatie? Volgens Hoogerwerf (1991) zijn de oorzaken voor te weinig kwaliteit van beleidsevaluaties:

- Beleidsmensen zijn vaak weinig of niet op de hoogte van de wetenschappelijke eisen die aan evaluatieonderzoek worden gesteld;
- Weinig politieke belangstelling voor evaluatie, politici zijn meer in het maken van plannen geïnteresseerd dan in het beoordelen van hun resultaten (ook uit angst);
- Evaluatieonderzoekers zijn vaak onvoldoende op de hoogte van de eisen die vanuit de beleidspraktijk aan het evaluatieonderzoek worden gesteld, waardoor het onderzoeksrapport daardoor veelal onvoldoende op het beleid is afgestemd;
- Het onderzoeksrapport staat vaak onder tijdsdruk en lijdt dikwijls onder een tekort aan personele en financiële middelen;
- Ontbreken van benodigde gegevens;
- Teveel politieke druk waardoor de evaluator niet in staat wordt gesteld om een objectieve evaluatie te verrichten.

Tot slot speelt er nog het belangrijke probleem dat te maken heeft met het onderzochte beleid. Het in kaart brengen van het **beleidsveld** is geen gemakkelijke opgave. Zo is het één en ander vaak ondoorzichtig, bijvoorbeeld: (1) processen gaan geleidelijk, vlakken in de uitvoering altijd wat af, lopen door elkaar heen, (2) het gekozen beleid interfereert met eerder en ander beleid, en/of (3) de (sociale) situatie verandert. Hierdoor kan de problematiek in kwestie verschuiven.

In de methode van onderzoek worden deze problemen meegenomen en is een poging gedaan om deze zoveel mogelijk weg te werken of te beperken.





## 4 Meta-evaluatie

### 4.1 Inleiding

Door verschillende onderzoekers is dus aangegeven waar de problemen liggen met betrekking tot beleidsevaluatieonderzoek, maar hoe is de kwaliteit van een evaluatieonderzoek zélf vast te stellen? Deze vraag staat min of meer centraal in dit onderzoek en zal ook moeten worden beantwoord. Hiertoe wordt dit hoofdstuk een poging gedaan. De in hoofdstuk 3 genoemde problemen spelen een belangrijke rol in het opstellen van een methode van onderzoek om kritisch te kunnen oordelen over de kwaliteit van beleidsevaluatieonderzoeken. Alvorens een methode van onderzoek op te stellen is het van belang een ‘nieuw’ begrip te introduceren, te weten meta-evaluatie. In dit hoofdstuk wordt dit begrip uitgewerkt en wordt toegelicht hoe dit in het verdere verloop van dit onderzoek zal worden gebruikt.

### 4.2 Meta-evaluatie

Bressers (2003) is van mening dat pas wanneer informatie in een evaluatieonderzoek “gekend, erkend en benut” wordt er sprake kan zijn van leren door evalueren. In andere woorden, dan pas krijgen de resultaten uit een evaluatieonderzoek meerwaarde.

Het onderzoek naar benutting van evaluatieonderzoek wordt ook wel meta-evaluatieonderzoek genoemd. Om evaluatieonderzoek te kunnen beoordelen moet een onderzoeksmethode worden gebruikt waarbij alle eerder genoemde aspecten, die bij een evaluatieonderzoek zich doen gelden, worden meegenomen. Michael Scriven was het in 1969 die het begrip ‘meta-evaluatie’ introduceerde (Stufflebeam, 1974). Hij bedoelde hiermee simpel gezegd een evaluatie van een evaluatie(onderzoek). Stufflebeam (1974) erkende ook het belang van meta-evaluatie: *“Good evaluation requires that evaluation efforts themselves be evaluated. Many things can and often do go wrong in evaluation work. Accordingly, it is necessary to check evaluations for problems such as bias, technical error, administrative difficulties, and misuse. Such checks are needed both to improve ongoing evaluation activities and to assess the merits of completed evaluation efforts.”* (p. 1).

Het doel van meta-evaluatie is volgens Hoogerwerf (1991) in eerste instantie de kwaliteit van een evaluatie(onderzoek) te beoordelen. Een verder reikend doel is daarbij doorgaans om door middel van een meta-evaluatie de kwaliteit van evaluaties te verbeteren. Nog weer een ander doel is veelal om door verbetering van de kwaliteit van evaluaties ook het gebruik van evaluatieonderzoek in het beleidsproces te bevorderen. Dit sluit aan bij de doelstelling zoals die voor dit onderzoek is geformuleerd.

Er gelden wel voorwaarden aan het opstellen van een methodiek voor meta-evaluatie. Omdat meta-evaluatie zelf ook een vorm is van evaluatie gelden hiervoor dezelfde eisen als aan een ‘normale’ evaluatie. Stufflebeam (1974) heeft acht voorwaarden opgesteld voor meta-evaluatie die afgeleid zijn van de eisen aan evaluatie:

1. Meta-evaluatie betekent het vaststellen van de waarde van een ander evaluatieonderzoek;
2. Meta-evaluatie moet proactief informatie leveren die gebruikt kan worden om de kwaliteit van toekomstige evaluaties te verbeteren, en retroactief om de evaluator verantwoordelijk te maken van zijn of haar gedane evaluatie;
3. Meta-evaluatie moet het belang van de evaluatiedoelen vaststellen, de juistheid van het evaluatieonderzoeksontwerp, de implementatie van het onderzoeksontwerp en de kwaliteit en het belang van de evaluatieresultaten;

4. Meta-evaluatie moet de evaluatie van onderzoek beschrijven en beoordelen én een aanbeveling bevatten evaluatie kan worden verbeterd en hoe de bevindingen in de toekomst kunnen worden gebruikt;
5. Meta-evaluatie moet in dienst staan van de evaluators en de alle personen die geïnteresseerd zijn in hun werk;
6. Evaluators zouden meta-evaluatie moeten uitvoeren. Zij moeten op hun beurt weer door externen beoordeeld worden op de kwaliteit van hun werk;
7. Meta-evaluatie moet uit drie stappen bestaan. Eerst moeten de specifieke meta-evaluatieonderzoeksvragen worden vermeld, vervolgens moet de benodigde data worden verzameld, gerangschikt en geanalyseerd worden. Tot slot moet de verkregen informatie worden toegepast op de besluitvorming.
8. Meta-evaluatie moet technische adequaat, nuttig, en kosteneffectief zijn.

Een belangrijk aandachtspunt bij de evaluatie van evaluatieonderzoek is de relatie tussen opdrachtgever en onderzoeker. Deze speelt een vooraanstaande rol bij kritiek op wetenschappelijk evaluatieonderzoek. In toenemende mate wordt vooraf overleg gevoerd over de precieze onderzoeksvraagstelling, de rapportage over het onderzoek en het gebruik van de resultaten. Dit raakt aan de rol van de onderzoeker en het thema van opvattingen over wetenschap (Stokking, 1991). Een onderzoeker is op deze manier immers niet meer onafhankelijk. Met het oog op wetenschappelijk onderzoek waaraan belangrijke conclusies en maatregelen zijn verbonden, moet dit zoveel mogelijk worden voorkomen. In de praktijk echter blijkt dit moeilijk realiseerbaar.

### 4.3 Meta-evaluatie-checklists

Eén van de methoden om een meta-evaluatie uit te voeren is ontwikkeld door Stufflebeam. Hij heeft de zogenaamde 'evaluation checklist' ontwikkeld. Deze checklist is een lijst om na te gaan of aan alle voorwaarden aan evaluatieonderzoek is voldaan. Stufflebeam (1974) kwam tot de conclusie data er geen eenduidige methoden voor het uitvoeren van meta-evaluatie voorhanden waren. De beschikbare middelen om meta-evaluatie uit te voeren waren volgens hem niet uitgebreid genoeg of niet voldoende georganiseerd, maar het belangrijkste was het gebrek aan gepubliceerde methoden voor meta-evaluatie. Daarom ontwikkelde Stufflebeam deze checklists zelf, naar aanleiding van vragen van zijn studenten naar een raamwerk voor het uitvoeren van evaluatieonderzoek. Shepard (1977) is van mening dat een checklist vele voordelen kent, omdat deze 'down-to-earth' en praktisch is. Een checklist verzekert de onderzoeker ervan dat hij alle potentiële categorieën meeneemt in zijn onderzoek, terwijl er geen tijd verloren gaat aan het onderzoek naar zaken die er niet toe doen.

Scriven (2000b) heeft hier nog enkele aanvullingen op, zoals:

- checklists zijn makkelijker te begrijpen voor de betrokken partijen;
- er wordt voorkomen dat één criterium teveel invloed heeft op de totale uitkomst van het onderzoek;
- 'dubbelweging' wordt uitgesloten;
- checklists zijn makkelijker op te stellen dan andere methoden van onderzoek.

Al deze punten maken een checklist uitermate geschikt voor dit onderzoek naar evaluatieonderzoek. Stufflebeam begon in 1974 met het opstellen van criteria waaraan evaluatieonderzoek zou moeten voldoen. Hierbij hanteerde hij vier criteria voor technische adequaatheid en zes gebruikscriteria. De vier criteria voor technische adequaatheid luiden als volgt:

1. Interne validiteit. Hierbij gaat het om de mate waarin de bevindingen waar zijn. Geeft de evaluatie antwoord op de vraag zoals bedoeld is? Zijn de resultaten accuraat en ondubbelzinnig?

2. Externe validiteit. Hierbij gaat het om de generaliseerbaarheid van de informatie. Kan de informatie ook voor andere doelgroepen worden gebruikt? Is het onderzoek tijdsgebonden of geeft het een voorspelling voor wat in de toekomst kan gebeuren?
3. Betrouwbaarheid. Dit heeft te maken met de nauwkeurigheid van de data. Zijn de data consistent?
4. Objectiviteit. Zijn de bevindingen wel objectief, of zijn ze door de evaluator (degene die de evaluatie uitvoert) beïnvloed?

Daarnaast geeft Stufflebeam zes gebruikscriteria met betrekking tot de bruikbaarheid waaraan evaluatierapporten moeten voldoen:

1. Relevantie. Hierbij gaat het om de vraag of de evaluatie voldoet aan de voorwaarden die gesteld worden.
2. Betekenis / belang. Dit gaat om de data die verzameld moeten worden. Niet alle data zijn namelijk relevant voor het onderzoek. Daarnaast is het onmogelijk om alle relevante data te verzamelen. Daarom is het van belang dat de evaluator zoveel mogelijk de data verzameld die het meest het doel dienen van het onderzoek<sup>12</sup>.  
*“To do this, the evaluator needs to rate the importance of each potentially relevant datum and he needs to know the priorities the audiences assign to the various data. Then, based on his own judgments and those of his audiences, the evaluator needs to choose the most significant data.”* (p.8).<sup>13</sup>
3. Scope. De informatie van de evaluatie moet het juiste bereik hebben. Sommige informatie is relevant en belangrijk, maar geeft niet alle antwoorden op vragen van het publiek.
4. Geloofwaardigheid. Vertrouwt het publiek de evaluator? Volgens Stufflebeam is dit een cruciaal onderdeel van een evaluatieonderzoek.
5. Tijdsaspect. Dit is erg belangrijk: de beste informatie is waardeloos wanneer deze informatie te laat komt om het doel van de evaluatie te dienen. In tegenstelling tot wetenschap (waarbij het gaat om kennis te ontwikkelen) is het doel van evaluatie om de uitvoering (van beleid) te veranderen. Daarom moeten de data op tijd worden geleverd. De tijdsdruk zorgt bijna altijd voor conflicten: als een evaluator alle informatie wil verzamelen die hij nodig heeft, dan is het rapport zeker niet klaar op het gevraagde moment. Daarom moet de evaluator volgens Stufflebeam streven om redelijke juiste informatie te verstrekken op het moment dat het ‘publiek’ daarom vraagt.
6. Doorwerking. Dit laatste criterium met betrekking tot de bruikbaarheid betreft de doorwerking van de resultaten. Aan dit criterium wordt voldaan wanneer alle personen waarvoor de evaluatie was bedoeld ook daadwerkelijk de resultaten krijgen én ze ook gebruiken.

Toch geven volgens Stufflebeam de vier technische en zes gebruikscriteria onvoldoende de moeilijkheid van de evaluatie aan. Het evaluatierapport moet worden beoordeeld op grond van wetenschappelijke criteria die gelden voor elk wetenschappelijk rapport. Maar het rapport moet ook worden beoordeeld op de aangegeven criteria.

Stufflebeam maakt het voor de evaluator nog moeilijker: er is nog een laatste standaard waaraan het evaluatieonderzoek moet voldoen, namelijk het *kosten- en effectiviteitsaspect*. Het gaat hierbij om het zo laag mogelijk houden van de kosten zonder de kwaliteit van de evaluatie aan te tasten. De uiteindelijke “Program Evaluations Metaevaluation Checklist” van Stufflebeam is opgenomen in de bijlage.

<sup>12</sup> Dit is lastig vast te stellen voor een buitenstaander, je weet immers als ‘leek’ niet welke data van belang zijn.

<sup>13</sup> Stufflebeam meent dus dat de evaluator de relevante data moet vaststellen, maar dit kan strijdig zijn met de eis van objectiviteit. De evaluator kan immers het onderzoek op een bepaalde manier sturen. Daarom is het van belang om voor dat de evaluator daadwerkelijk met het onderzoek begint, in overleg af te stemmen welke data relevant zijn.

Ook andere wetenschappers dan Stufflebeam hebben zich bezig gehouden met het ontwikkelen van een methode voor meta-evaluatie. In een schema zoals is opgesteld door Hoogerwerf et al. (1991) zijn enkele eisen geformuleerd ten aanzien van meta-evaluatie van een beleidsevaluatieonderzoek. Uit dit schema blijkt echter ook weer het terugkerende probleem: het is moeilijk aan te geven waar een *evaluatie*onderzoek(rapport) aan zou moeten voldoen. Het schema geeft in onvoldoende mate aan wat specifiek het onderscheid is tussen een evaluatieonderzoek en (wetenschappelijk) onderzoek in bredere zin; enkele genoemde eisen zijn te algemeen. Dat wil zeggen dat ze aangeven waar een rapport in ieder geval aan moet voldoen, maar dat ze te weinig benadrukken waar een goed evaluatierapport aan moet voldoen. Uit het schema kunnen enkele relevante eisen aan het onderzoeksrapport worden gehaald die wel specifiek betrekking hebben op de kwaliteit van een evaluatierapport en niet op rapporten in het algemeen. De relevantie van de eisen hangt af van de mate waarin ze onderscheidend zijn tussen 'normaal' onderzoek en daadwerkelijk evaluatieonderzoek, oftewel algemene eisen, die voor elk wetenschappelijk onderzoek gelden, worden hier niet genoemd. De verkorte lijst met criteria, die volgens Hoogerwerf voor evaluatieonderzoeksrapporten gelden, is als volgt:

- Het rapport moet inzicht geven in de doelbereiking, effectiviteit, neveneffecten en efficiëntie van het beleid;
- In het rapport moet worden vermeld en beargumenteerd of het onderzoek verklarend of beschrijvend is;
- Het rapport bevat één of meerdere duidelijk geformuleerde en beargumenteerde maatstaven;
- Wijze waarop data zijn geanalyseerd is in het rapport duidelijke omschreven en beargumenteerd;
- De wijze waarop de geanalyseerde data wordt geïnterpreteerd is expliciet en duidelijk omschreven in het rapport;
- Het rapport verwijst naar één of meer bepaalde beleidsdoeleinden van bepaalde personen of organisaties en gaat na of deze al dan niet dankzij het beleid zijn bereikt of kunnen worden bereikt;
- Het rapport gaat na wat het beleid tot de oplossing van het probleem heeft bijgedragen of kan bijdragen;
- Het rapport geeft een duidelijke beoordeling aan de hand van één of meer expliciet en duidelijk vermelde maatstaven (ontleend aan beginselen, normen of doeleinden) uit de beleidspraktijk;
- Het rapport verwijst, indien het een effectiviteitsonderzoek betreft, naar één of meer bepaalde beleidsinstrumenten (middelen) en gaat na in hoeverre deze hebben bijgedragen of kunnen bijdragen tot het bereiken van het doel waarvoor zij worden ingezet;
- Het rapport besteedt aandacht aan neveneffecten van het beleid;
- Het rapport maakt t.a.v. de effecten van het beleid onderscheid tussen positief gewaardeerde effecten (baten of voordelen) en negatief gewaardeerde effecten (kosten of nadelen) en tracht deze met elkaar te vergelijken;
- Het rapport verwijst naar eventuele tijdskeuzen (t.a.v. tijdstippen, tijdsindeling, volgorde, tijdsduur of tempo) van het beleid en gaat na in hoeverre het beleidsproces of de beleidseffecten in overeenstemming zijn met de gedane tijdskeuzen;
- Het rapport geeft ook in zijn conclusies een beoordeling van (de effecten van) een gevoerd of te voeren beleid (in termen van waarden, normen en doeleinden);
- Het rapport geeft ten aanzien van verklarende variabelen - waaronder de beleidsuitvoering en bepaalde kenmerken van het beleidsveld - aan in hoeverre deze vatbaar zijn voor beïnvloeding door de beleidsinstantie, dus manipuleerbaar zijn.

Wanneer kritisch gekeken wordt naar de eisen die Hoogerwerf stelt, dan blijken zij in grote mate overeen te komen met de eisen uit de evaluatie-checklist van Stufflebeam. De checklist Stufflebeam en de criteria van Hoogerwerf zijn in het volgende hoofdstuk als basis genomen voor het ontwikkelen van een methode van onderzoek.

Tot slot nog dit: onderzoek naar evaluatieonderzoek heeft volgens Bressers (2003) geleid tot de conclusie dat ook als niet meteen gebruik wordt gemaakt van een onderzoek, de resultaten alsnog wel ergens kunnen blijven hangen en in een later stadium worden gebruikt. Dit zal wel leiden tot een verbetering van de algemene opvattingen en zo de kwaliteit van beleid verbeteren. Met andere woorden, om het meta-evaluatieonderzoek succesvol te laten zijn hoeven de conclusies uit het onderzoek niet direct te worden vertaald naar aanpassing van het beleid, als het maar 'ergens blijft hangen'. Voor de beoordeling van een evaluatieonderzoek is dit een complicerende factor. Het evaluatieonderzoek kan, wanneer het wordt beoordeeld aan de hand van de criteria die gelden voor evaluatieonderzoek, als ondermaats worden beoordeeld, maar tegelijkertijd wel een discussie op gang brengen waardoor uiteindelijk de kwaliteit van het beleid wel in belangrijke mate wordt verbeterd. Dit is echter zeer moeilijk te meten en wordt, mede gelet op het tijdsaspect, in dit onderzoek naar evaluatieonderzoek niet meegenomen.



## 5 Methode van onderzoek

In dit hoofdstuk is de methode van onderzoek uitgewerkt. De methode van onderzoek is goed beschouwd al begonnen bij het literatuuronderzoek om beleid, evaluatie en evaluatieonderzoek te beschrijven. Dit is echter niet voldoende om te komen tot een methodiek waarmee evaluatieonderzoeken zelf onderwerp van evaluatie kunnen worden gemaakt. Hiervoor is het nodig de verzamelde informatie te rangschikken en te ordenen. Dat is in dit hoofdstuk eerst uitgewerkt. Vervolgens worden de methoden die voor het uitvoeren van meta-evaluatie worden gebruikt beschreven. Tot slot wordt een onderzoeksmethode ontwikkeld, waarmee beleidsevaluaties kunnen worden beoordeeld op hun kwaliteit, aan de hand van uit de literatuur verzamelde criteria.

Allereerst kan gesteld worden dat de methode van onderzoek gebaseerd is op de methoden voor ex post evaluatie. Evaluatieonderzoek ex post kan gedefinieerd worden als: *“systematisch onderzoek naar de effecten van bestaand beleid, de wijze waarop beleid wordt uitgevoerd en/of de kosten en de kwaliteit van de geleverde producten of diensten”* (Ministerie van Financiën, 2003, p.9).

Bij ex post evaluatieonderzoek wordt uitgegaan van informatie die al beschikbaar is, waarop het oordeel van doeltreffendheid en doelmatigheid is gebaseerd. Door alle verzamelde informatie naast elkaar te leggen kan een methodiek worden ontwikkeld voor het vaststellen van de kwaliteit van een evaluatie.

De vragen die in de casestudie voor dit onderzoek worden beantwoord dienen om de centrale vraag bij evaluatieonderzoek te beantwoorden: **het rapport moet inzicht geven in de doelbereiking, effectiviteit, neveneffecten en efficiëntie van het beleid.**

De uiteindelijke methodiek voor het bepalen van de kwaliteit van een beleidsevaluatieonderzoek of -rapport is niet gericht op de eisen waaraan elk onderzoek zou moeten voldoen. Deze zijn niet bepalend voor de mate waarin gesproken mag worden van evaluatieonderzoek. De eisen die voor dit onderzoek zijn geformuleerd zijn afgeleid en een samenvatting van de meta-evaluatie ‘checklists’ van Stufflebeam en van de eisen die Hoogerwerf en Arts stellen aan evaluatieonderzoek. Door de checklist van Stufflebeam en de criteria voor meta-evaluatie van Hoogerwerf te combineren kan een lijst met criteria worden samengesteld die de kwaliteit van een evaluatieonderzoek aangeeft. De lijst met criteria is echter zo uitgebreid dat het overzicht verloren dreigt te gaan. Daarom worden deze criteria bij de methode van onderzoek gerangschikt. Hiervoor is gebruik gemaakt van het onderscheid dat door Arts (1998) is gemaakt tussen monitoring, auditing en evaluatie. Deze indeling wordt als uitgangspunt genomen omdat deze inzicht geeft in de mate van ‘moeilijkheid’ van het evaluatieonderzoek.

Eisen aan beleid	Onderzoekstype Bressers	Onderzoekstype Arts
Doelbereiking	Doelbereikingsonderzoek	Monitoring
Doeltreffendheid	Effectiviteitsonderzoek	Auditing
Doelmatigheid	Efficiëntieonderzoek	Evaluatie
Legitimiteit	X	X

**Figuur 5.1** Onderscheid tussen evaluatieonderszoekstypen



Dit onderscheid betekent niet dat wordt afgeweken van het onderscheid dat eerder werd gemaakt, tussen doelbereiking, effectiviteit en efficiëntie. Arts onderkent dit onderscheid namelijk ook en neemt dit ook op. Monitoring is een waarneming, auditing geeft inzicht in doelbereiking en evaluatie handelt over de effectiviteit en efficiëntie van beleid (houdt dus ook monitoring en auditing). Hij plaats echter hier een kritische kanttekening bij. De genoemde eisen aan beleid zijn volgens hem wel belangrijk, maar hij ziet ook doorwerking van beleid als één van de belangrijkste criteria voor beleid. Daarom zou dit ook moeten worden gemeten, alhoewel hij ook aangeeft dat doorwerking lastig te meten is.

Het is verder opvallend dat zowel Hoogerwerf als Stufflebeam de eis van legitimiteit van een beleid niet laten terugkomen. Hoewel de legitimiteit voor een deel besloten ligt in vooral de efficiëntie van beleid, is de mate waarin wordt voldaan aan deze eis wel van belang. Daarom is naast de eisen van Hoogerwerf en Stufflebeam ook de eis opgenomen dat het rapport inzicht moet geven in de legitimiteit van het beleid.

Op basis van de lijst met criteria<sup>14</sup> kan bepaald worden of het een monitorings-, auditing- of evaluatierapport betreft. Bij evaluatierapport kan weer onderscheid worden gemaakt tussen een effectiviteitsonderzoek, efficiëntieonderzoek en wetenschappelijk evaluatieonderzoek.

Er bestaat onderscheid tussen wetenschappelijke eisen aan evaluatieonderzoek en eisen die worden gesteld uit de beleidspraktijk. Vooral Hoogerwerf (1991) heeft hierop gewezen. Evaluatieonderzoek is wetenschappelijk onderzoek, dus zijn wetenschappelijke eisen ter zake. Maar het is ook onderzoek naar beleid en veelal tevens onderzoek ten behoeve van beleid, dus zijn ook eisen vanuit de beleidspraktijk relevant. Deze tweedeling hoeft allerm minst een onverzoenlijke tegenstelling op te leveren. De wetenschap en het (openbaar) bestuur zijn weliswaar twee verschillende maatschappelijke systemen, elk met hun eigen structuur, cultuur en processen. Zij raken elkaar echter op allerlei wijzen, voeren een kritische dialoog en kunnen ook, met behoud van hun onafhankelijkheid, elkaar op vele manieren tot steun zijn<sup>15</sup>.

In dit onderzoek wordt dit onderscheid wel onderkend, maar de grens is niet duidelijk aan te geven. Daarom wordt dit onderscheid hier niet gemaakt, maar wordt een lijst opgesteld van algemene eisen aan evaluatieonderzoek (dus zowel wetenschappelijke als uit de beleidspraktijk).

Al deze thema's in ogenschouw nemend, is gekomen tot deze lijst voor het beoordelen van het soort onderzoek en de kwaliteit hiervan. Hierbij is, als gezegd, onderscheid gemaakt tussen monitoring, auditing, evaluatie en algemene wetenschappelijke eisen. De grens tussen monitoring, auditing en evaluatie is, zo is gebleken, niet altijd eenvoudig aan te geven. Vaak overlappen de criteria elkaar gedeeltelijk, waardoor het onduidelijk blijft om welk soort onderzoek het gaat.

---

<sup>14</sup> Een kanttekening is hierbij op zijn plaats: de lijst is aan veranderingen onderhevig en is nooit 'af' en 'compleet'. Het feit dat Stufflebeam al sinds het begin van de jaren zeventig sleutelt aan zijn 'checklist' en nog steeds niet klaar is, ontnemt de illusie dat in dit onderzoek gehanteerd checklist compleet is. Toch kan het een goede en snelle methode zijn om de kwaliteit van een evaluatierapport tot een bepaalde hoogte vast te stellen. Daarnaast reikt het voor dit onderzoek te ver om bij de beoordeling in te gaan op de oorzaak van de eventueel gebrekkige kwaliteit van de evaluatie. Hier zijn veel verschillende factoren van invloed, waarvan er enkele elders in dit rapport zijn genoemd. In de uitgewerkte casestudies is er te weinig zicht op de relaties en processen die aan de evaluatie ten grondslag liggen. Dit onderzoek is daarom beperkt tot het bepalen van de kwaliteit van de evaluatieonderzoek.

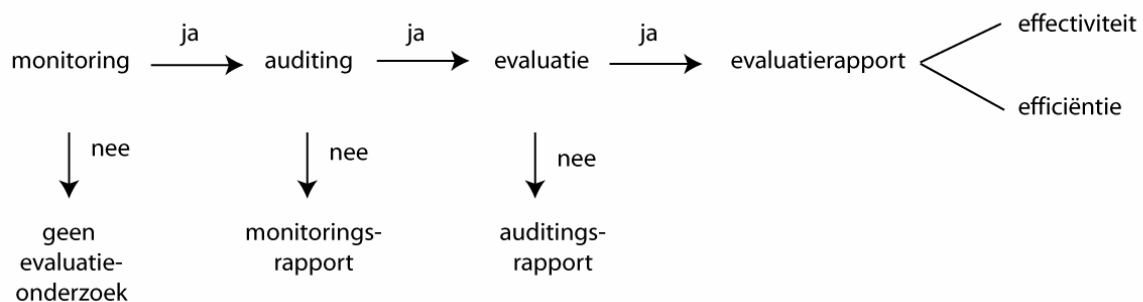
<sup>15</sup> Toch bestaat er ook een zekere spanning. Zo heeft de beleidspraktijk vaak behoefte aan een snelle evaluatie, kort na de start van de beleidsuitvoering, terwijl de wetenschappelijke onderzoeker weet dat er enige tijd moet verlopen alvorens de effecten van het beleid kunnen zijn opgetreden en dus zinvol kunnen worden gemeten. Bovendien kost goed wetenschappelijk onderzoek en dus ook goed evaluatieonderzoek zelf doorgaans veel tijd en geld.

<p><b>MONITORING</b></p> <ul style="list-style-type: none"> <li>† Het rapport geeft ontwikkelingen aan uit de beleidspraktijk door middel van metingen en observaties van relevante variabelen over een bepaald tijdsbestek en met een bepaald doel</li> </ul>
<p><b>AUDITING</b></p> <ul style="list-style-type: none"> <li>† Het rapport bevat één of meerdere duidelijk geformuleerde en beargumenteerde maatstaven en geeft een duidelijke beoordeling aan de hand van deze vermelde maatstaven (doeleinden) uit de beleidspraktijk</li> <li>† Het rapport geeft inzicht in de mate van doelbereiking</li> <li>† Het rapport maakt ten aanzien van de effecten onderscheid tussen positief gewaardeerde effecten (baten of voordelen) en negatief gewaardeerde effecten (kosten of nadelen) en tracht deze met elkaar te vergelijken</li> <li>† Het rapport beesteedt aandacht aan de neveneffecten van het beleid en geeft aan of deze negatief of positief worden gewaardeerd</li> </ul>
<p><b>EVALUATIE</b></p> <ul style="list-style-type: none"> <li>† Het rapport verwijst naar één of meer bepaalde beleidsdoeleinden van bepaalde personen of organisaties en gaat na welke uitwerking het beleid uiteindelijk heeft gehad of kan hebben</li> <li>† Het rapport geeft een verklaring voor het al dan bereiken van de beleidsdoelstellingen</li> <li>† Om een verklaring te kunnen geven is het nodig inzicht te hebben in de relaties in het beleidsveld. Het rapport moet daarom een beschrijving geven van het beleidsveld en de daarin optredende relaties</li> <li>† Het rapport gaat na wat het beleid tot de oplossing van het probleem heeft bijgedragen of kan bijdragen (vereist een duidelijke probleemformulering)</li> <li>† Het rapport verwijst naar één of meer bepaalde beleidsinstrumenten en gaat na in hoeverre deze hebben bijgedragen of kunnen bijdragen aan het bereiken van het doel waarvoor zij worden ingezet</li> <li>† Het rapport moet inzicht geven in de interveniërende variabelen (externe invloeden), die van invloed kunnen op het al dan niet bereiken van de beleidsdoelstellingen</li> <li>† Het rapport geeft ten aanzien van verklarende variabelen (waaronder de beleidsuitvoering en bepaalde kenmerken van het beleidsveld) aan in hoeverre deze vatbaar zijn door beïnvloeding door beleidsinstantie, dus manipuleerbaar zijn</li> <li>† Het rapport moet inzicht geven in de sterke en zwakke punten van het beleid</li> <li>† Het rapport moet aangeven in hoeverre de sterke punten van het beleid kunnen worden gebruikt om zwakke punten weg te werken of te beperken</li> <li>† <i>Het rapport moet inzicht geven in de mate waarin het geëvalueerde beleid wordt gedragen door de betrokken partijen</i></li> </ul>
<p><b>ALGEMEEN WETENSCHAPPELIJK</b></p> <ul style="list-style-type: none"> <li>† De wijze waarop de data zijn beoordeeld, geanalyseerd en geïnterpreteerd is duidelijk en expliciet omschreven in het rapport</li> <li>† In het rapport moet worden aangegeven welke beperkingen het onderzoek kent, met betrekking tot dataverzameling, methode, referentiemateriaal, etcetera met het oog op de geldigheid van de conclusies</li> <li>† Het rapport moet gefocust zijn op de belangrijkste vragen ('sleutelvragen') en de conclusies moeten in de eerste plaats op deze vragen gericht zijn</li> <li>† In de conclusie(s) van het rapport moet duidelijk worden aangegeven voor welk tijdsbestek, omstandigheden, doel en activiteiten het onderzoek geldt</li> </ul>

**Figuur 5.2** Criteria voor evaluatieonderzoek

De grote pijl aan de linkerkant geeft aan dat de lijst met criteria voor monitoring, auditing en evaluatie in aflopende volgorde wordt doorlopen. De algemene wetenschappelijke eisen gelden voor het rapport op zich en kan onderscheidend zijn wanneer aan alle eisen is voldaan. De kleuren zijn toegevoegd om bij de beoordeling duidelijk te kunnen maken welk soort onderzoek het betreft.

Op basis van de lijst is daarnaast een simpele, maar duidelijke, beslisboom ontwikkeld. Het heeft immers weinig zin om de checklist verder te doorlopen, wanneer is vastgesteld dat aan de eerste voorwaarde onder monitoring al niet wordt voldaan. Het rapport zal in dit geval zeker geen evaluatierapport zijn. De minimumeis is daarom ook dat het rapport tenminste een bepaalde mate van monitoring moet inhouden.



**Figuur 5.3 Beslisboom evaluatieonderzoek**

Indien het onderzoeksrapport een evaluatierapport betreft, kan onderscheid gemaakt worden tussen de verschillende vormen van evaluatieonderzoek, en de mate waarin het rapport inzicht geeft in doelbereiking, effectiviteit (doeltreffendheid) en efficiëntie (doelmatigheid).

De eis aan inzicht in legitimiteit kan min of meer als op zichzelf staand worden beschouwd. Hoewel legitimiteit als eis wordt gesteld aan beleid, hoeft nog niet te betekenen dat wanneer niet voldaan wordt aan de eisen die gelden voor evaluatie, er ook geen inzicht wordt gegeven in de legitimiteit van het beleid. Dit betekent voor de methode van onderzoek dat de eis van legitimiteit altijd wordt beoordeeld, ook al is het onderzoek niet meer dan een monitoringsrapport.

Op basis van de vergelijking van de onderzoeken is achteraf aangegeven wat de score is van de onderzoeken op ieder criteria (voorzover mogelijk). Bij het toekennen van de scores is uitgegaan van een driedeling. Een plus-score wordt gegeven als het punt uit de checklist volledig voldoet aan de gestelde eis. Een neutrale score wordt gegeven wanneer slechts gedeeltelijk wordt voldaan aan eis of wanneer de uitwerking onvoldoende in het rapport terugkomt. Een min-score wordt gegeven wanneer niet wordt voldaan aan de eis.

-	Voldoet niet aan dit criterium
0	Voldoet gedeeltelijk en/of uitwerking is niet voldoende
+	Voldoet aan criterium

Elk onderzoek dat in hoofdstuk 6 wordt genoemd, wordt beoordeeld aan de hand van de 19 eisen die zijn geformuleerd op basis van deze schaalverdeling. Dit gebeurt aan de hand van een tabel met de volgende indeling:

Criteria	Beoordeling			Opmerkingen
xxxxx	+	0	-	xxxxx

De beoordeling wordt gegeven aan de hand van +/0/- én met een kleur (+ = lichtgrijs, 0 = grijs, - = donkergrijs), om de beoordeling overzichtelijker te maken. Indien nodig wordt in de kolom 'opmerkingen' een nadere toelichting gegeven op de beoordeling.

Daarnaast wordt gekeken wie de opdrachtgever is en door wie het onderzoek vervolgens wordt uitgevoerd. Wellicht zijn hier, in relatie tot de criteria, conclusies aan te verbinden met betrekking tot de kwaliteit van de uitgevoerde evaluatie.

Tot slot wordt voor elk onderzoek afzonderlijk conclusies getrokken. Ook worden algemeen geldende conclusies getrokken.



## DEEL II – DE PRAKTIJK

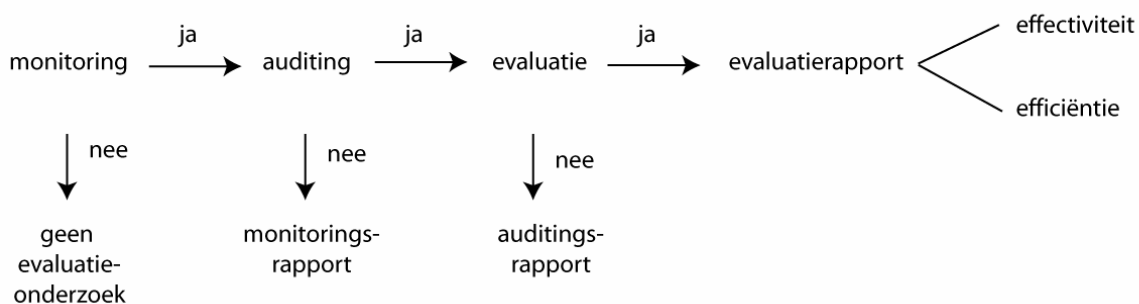
### 6 Casestudies

#### 6.1 Inleiding

In dit hoofdstuk zijn enkele evaluatieonderzoeken geëvalueerd aan de hand van de lijst met criteria zoals is opgesteld in hoofdstuk 5. Het doel hiervan is inzicht te krijgen in de kwaliteit van onderzoeken uit de praktijk én om aanbevelingen te kunnen doen over aspecten die beter kunnen en extra aandacht vereisen. Daarnaast wordt gekeken naar algemeen wetenschappelijke eisen die voor elk onderzoek gelden.

Het is echter zo dat de hier genoemde onderzoeken slechts een zeer klein deel van het totaal aantal evaluatieonderzoeken beslaat die de laatste jaren zijn uitgevoerd. De conclusies gelden daarom ook slechts voor de onderzochte rapporten. Wel kan aan de hand van de beoordeling van deze rapporten inzicht worden verkregen in de aspecten die extra aandacht vereisen en waar in de praktijk overheen wordt gekeken. Dit kan een bijdrage leveren aan de kwaliteit van toekomstige evaluatieonderzoeken en aan de kwaliteit van beleid.

De werkwijze dit gevolgd wordt is zoals die in de eerder opgestelde beslisboom wordt voorgesteld:



**Figuur 6.1 Beslisboom evaluatie**

Deze beslisboom vergemakkelijkt het doorlopen van de stappen, aangezien bij elke stap vastgesteld kan worden of het evaluatieonderzoek voldoet aan de geldende criteria die voor de typen evaluatieonderzoek gelden. Wanneer niet voldaan aan de criteria, dan kan volstaan worden met het vaststellen dat het respectievelijk geen evaluatieonderzoek, een monitorings- / effectenrapport of een auditings- / doelbereikingsrapport betreft. Hiermee wordt voorkomen dat onnodig aan extra criteria getoetst hoeft te worden<sup>16</sup>. Voor elk type onderzoek wordt wel gekeken naar de algemeen geldende wetenschappelijke eisen.

Er is gekozen om de doelstellingen zoals die in de evaluatieonderzoeken zijn opgenomen, niet stuk voor stuk bij langs te lopen, aangezien dit niks toevoegt aan de methodiek van dit onderzoek. Doel is immers niet om het beleid op zich te beoordelen, maar de uitvoering van de beleidsevaluatie.

<sup>16</sup> De rapporten worden slechts beoordeeld op de vastgestelde criteria. Er wordt dus geen onderzoek gedaan in hoeverre de uitspraken die in de evaluatieonderzoeken worden gedaan, ook daadwerkelijk juist en waar zijn. Dit is onmogelijk te onderzoeken in een relatief kort tijdsbestek.

## 6.2 Beleidseffectrapportage Verkeer & Vervoer Noord-Nederland, periode 1986-2001

Het eerste rapport dat is geëvalueerd is de "Beleidseffectrapportage Verkeer & Vervoer Noord-Nederland, periode 1986-2001". Deze is opgesteld door Grontmij Drenthe, in opdracht van het Samenwerkingsverband Noord-Nederland en Rijkswaterstaat Directie Noord-Nederland. De opdracht is dus extern uitgevoerd.

De subtitel van dit rapport luidt: *een overzicht van de effecten van het beleid in Noord-Nederland*. Dit is een misleidende titel, aangezien de doelstelling van het rapport is: "...inzicht te geven in de ontwikkelingen in de verkeers- en vervoerssector in Noord-Nederland". Dit is niet het doel van een beleidseffectrapportage. Hier zou 'het in kaart brengen van beleidseffecten' een betere beschrijving vormen, en eventueel het verklaren van het al dan niet optreden van deze effecten.

De beoordeling van het rapport aan de hand van de criteria uit het voorgaande hoofdstuk kan als volgt worden beschreven.

### Monitoring

criterium	Beoordeling	Opmerkingen
Het rapport geeft ontwikkelingen aan uit de beleidspraktijk door middel van metingen en observaties van relevante variabelen over een bepaald tijdsbestek en met een bepaald doel	+	Aan dit criterium wordt voldaan, de beoordeling hiervan is een +. Met andere woorden: er kan gesproken worden van een monitoringsrapportage.

### Auditing

criterium	Beoordeling	Opmerkingen
Het rapport bevat één of meer duidelijk geformuleerde en beargumenteerde maatstaven (normen, doeleinden) en geeft een duidelijke beoordeling aan de van deze vermelde maatstaven	-	Hier wordt niet aan voldaan. Afgaand op de opgestelde beslisboom heeft het geen zin om verder te gaan met het beoordelen aan de hand van de auditing- en evaluatiecriteria, aangezien de basis hiervoor ontbreekt.

De conclusie luidt dan ook dat dit slechts een monitoringsrapportage is.

### Legitimiteit

Het beleid moet inzicht geven in de mate waarin het beleid wordt gedragen door betrokken partijen.	-	
--	---	--

Beoordeling aan de hand van algemene wetenschappelijke eisen is nog wel zinvol, hierdoor kan de algemene kwaliteit van het rapport worden bepaald.

### Algemeen wetenschappelijk

criterium	Beoordeling	Opmerkingen
De wijze waarop de data zijn beoordeeld, geanalyseerd en geïnterpreteerd is duidelijk en expliciet omschreven in het rapport	-	
In het rapport moet aangegeven worden welke beperkingen het onderzoek kent, met betrekking tot dataverzameling, methode, referentiemateriaal, etcetera met het oog op de geldigheid van de conclusies	-	Geldigheid van conclusies is niet aan de orde, aangezien er überhaupt geen conclusies worden getrokken
Het rapport moet gefocust zijn op de belangrijkste vragen ('sleutelvragen') en de conclusies moeten in de eerste plaats op deze vragen gericht zijn	-	Idem
In de conclusie(s) van het rapport moet duidelijk worden aangegeven voor welk tijdsbestek, omstandigheden, doel en activiteiten het onderzoek geldt.	-	Idem

De conclusie aan de hand van bovengestelde criteria luidt dat wetenschappelijk gezien niet gesproken mag worden van een (evaluatie)onderzoek.

### Conclusie

Aan de hand van een beoordeling aan de hand van bovenstaande criteria kan de conclusie niet anders luiden dan dat dit **geen** beleidseffectrapportage is, laat staan een evaluatieonderzoek. Op geen enkele wijze wordt inzicht gegeven in het gevoerde beleid (met betrekking tot verkeer en vervoer) en de daaruit voortvloeiende effecten. Hoewel de relaties in de verkeers- en vervoerssector zeer complex zijn, schiet dit rapport tekort. Aangezien er wel verschillende ontwikkelingen worden gerapporteerd kan toch gesproken worden van een monitoringsrapportage.

De titel is misleidend en zou eigenlijk veranderd moeten worden in monitoringsrapportage<sup>17</sup>. Daarnaast is de aanbeveling om toch tot een bepaalde hoogte conclusies te trekken op basis van de indicatoren uit het rapport. Hierdoor kan in de toekomst het rapport verbeterd worden, en kan ook aan de hand van de conclusies worden vergeleken met voorgaande edities. Hierdoor zal het onderzoek kunnen opschuiven richting een auditing, en zal ook de kwaliteit van het verkeers- en vervoersbeleid in Noord-Nederland toenemen (mits lering wordt getrokken uit de evaluatie).

### 6.3 Beleidseffectrapportage Verkeer en Vervoer 1999

De Beleidseffectrapportage Verkeer en Vervoer 1999, met als ondertitel “beleidseffectmeting Verkeer en Vervoer”, is het volgende rapport dat is geëvalueerd. Dit rapport wordt jaarlijks in opdracht van Rijkswaterstaat opgesteld, extern uitgevoerd, en heeft als functie “een hulpmiddel zijn bij verschillende fasen in het beleids- en begrotingsproces, maar ook een bron van informatie en om het te kunnen verantwoorden. Voorts is het een hulpmiddel bij de communicatie over het verkeers- en vervoersbeleid zowel binnen het ministerie als met partners daarbuiten”.

Om aan deze functie te voldoen is een schema opgenomen waarin de plaats van het beleidseffectrapport in het beleidsproces wordt weergegeven. Dit schema is echter te algemeen en maakt niet duidelijk welke relaties er in de verkeers- en vervoerssector spelen. Het rapport veronderstelt dus dat effecten van het verkeers- en vervoersbeleid in beeld worden gebracht, en dat daarbij de rol die het beleid heeft gespeeld ook wordt uitgewerkt. Aan de hand van de criteria kan beoordeeld worden in hoeverre het rapport aan deze verwachting voldoet.

#### Monitoring

criterium	Beoordeling	Opmerkingen
Het rapport geeft ontwikkelingen aan uit de beleidspraktijk door middel van metingen en observaties van relevante variabelen over een bepaald tijdsbestek en met een bepaald doel	+	Aan dit criterium wordt voldaan, de beoordeling hiervan is een +. Met andere woorden: er kan gesproken worden van een monitoringsrapportage.

#### Auditing

criterium	Beoordeling	Opmerkingen
Het rapport bevat één of meer duidelijk geformuleerde en beargumenteerde maatstaven en geeft een duidelijke beoordeling aan de hand van deze vermelde maatstaven	+	Hier wordt aan voldaan, er zijn verschillende beleidsdoelstellingen opgenomen
Het rapport geeft inzicht in de mate van doelbereiking	+	Hier wordt op een duidelijke manier aan voldaan, het

<sup>17</sup> Aangezien ik betrokken ben geweest bij het opstellen van dit rapport heb ik de aanbeveling gedaan om het niet langer een beleidseffectrapportage te noemen (aangezien dit het niet is), maar een monitor verkeer en vervoer. Dit dekt beter de lading, aangezien deze monitor het volgen van ontwikkelingen of trends in de sector inhoudt.



		(tussen)doel wordt d.m.v. een grafiek aangegeven en de mate van doelbereiking wordt hier tegen afgezet
Het rapport maakt ten aanzien van de effecten onderscheid tussen positief gewaardeerde effecten (baten of voordelen) en negatief gewaardeerde effecten (kosten of nadelen) en tracht deze met elkaar te vergelijken.	0	Er wordt wel onderscheid gemaakt tussen positieve en negatieve effecten, maar deze worden niet met elkaar vergeleken
Het rapport besteedt aandacht aan de neveneffecten van het beleid en geeft aan of deze negatief of positief worden gewaardeerd	-	

De conclusie naar aanleiding van de auditing-criteria is dat het rapport hierop (licht) positief beoordeeld kan worden. Vooral het inzicht in neveneffecten kan beter.

### Evaluatie

Criterion	Beoordeling	Opmerkingen
Het rapport verwijst naar één of meer bepaalde beleidsdoeleinden van bepaalde personen of organisaties en gaat na welke de uitwerking het beleid uiteindelijk heeft gehad of kan hebben	0	Er worden wel beleidsdoeleinden opgenomen en het beleid wordt ook wel beschreven, maar de uiteindelijke invloed van het beleid blijft onderbelicht
Het rapport geeft een verklaring voor het al dan niet bereiken van de beleidsdoelstellingen	0	Een verklaring wordt wel gegeven, maar niet zozeer over het bereiken van de doelstellingen, als wel over 'ontwikkelingen'. Daarom is de beoordeling slechts 0. Ook is dit niet bij elke indicator het geval
Het rapport geeft een beschrijving van het beleidsveld en de daarin optredende relaties	-	
Het rapport gaat na wat het beleid tot de oplossing van het probleem heeft bijgedragen of kan bijdragen (duidelijke probleemstelling ook nodig)	0	Slechts tot op een bepaalde hoogte. Het beleid wordt wel genoemd, maar niet zozeer de invloed hiervan
Het rapport verwijst naar één of meer bepaalde beleidsinstrumenten en gaat na in hoeverre deze hebben bijgedragen of kunnen bijdragen aan het bereiken van het doel waarvoor zij zijn ingezet	0	Hiertoe wordt een poging gedaan, maar het wordt niet echt duidelijk wat de daadwerkelijke bijdrage van de beleidsinstrumenten is
Het rapport moet inzicht geven in de interveniërende variabelen (externe invloeden), die van invloed kunnen zijn op het al dan niet bereiken van de beleidsdoelstellingen	0	Externe invloeden worden wel genoemd, maar er wordt onvoldoende aangegeven wat de invloed hiervan op het proces is
Het rapport geeft ten aanzien van de verklarende variabelen (beleidsuitvoering en bepaalde kenmerken van het beleidsveld) aan in hoeverre deze vatbaar zijn voor beïnvloeding door de beleidsinstantie, dus manipuleerbaar zijn	-	Er wordt niet duidelijk hoe de maatregelen zijn geïmplementeerd en hoe deze door de beleidsinstantie kunnen worden beïnvloed
Het rapport moet inzicht geven in de sterke en zwakke punten van het beleid	0	Deze komen wel naar voren, maar worden niet expliciet genoemd. Het is meer toevallig dat deze duidelijk worden
Het rapport moet aangeven in hoeverre de sterke punten van het beleid kunnen worden gebruikt om de zwakke punten te beperken of weg te werken	-	
Het beleid moet inzicht geven in de mate waarin het beleid wordt gedragen door betrokken partijen.	0	Het draagvlak onder de bevolking wordt gemeten, maar niet onder betrokken partijen

De conclusie luidt dat dit rapport tekortschiet om een volwaardige evaluatierapport te mogen heten. Wel is een goede aanzet gegeven om in het vervolg wel aan de eisen van evaluatie te voldoen. De basis hiervoor is gelegd. Alle criteria vereisen hierbij de aandacht, maar extra aandacht moet uitgaan naar de beschrijving van het beleidsveld, waardoor ook de

implementatie en beïnvloeding van het beleid door de beleidsinstantie in kaart kan worden gebracht, en de sterke punten van het beleid tegenover de zwakke punten.

### Algemeen wetenschappelijk

criterium	Beoordeling	Opmerkingen
De wijze waarop de data zijn beoordeeld, geanalyseerd en geïnterpreteerd is duidelijk en expliciet omschreven in het rapport	0	Nog niet voldoende (niet bij elke indicator bijvoorbeeld)
In het rapport moet aangegeven worden welke beperkingen het onderzoek kent, met het betrekking tot de dataverzameling, methode, referentiemateriaal, etc. met het oog op de geldigheid van de conclusies	0	Niet altijd even goed duidelijk, hoewel er wel een poging toe wordt gedaan
Het rapport moet gefocust zijn op de belangrijkste vragen (sleutelvragen) en de conclusies moeten in de eerste plaats hierop gericht zijn	+	Wordt aan voldaan, alleen het gevaar bestaat wel dat er teveel andere indicatoren worden opgenomen, waardoor de aandacht wordt afgeleid van de sleutelvragen'
In de conclusie(s) van het rapport moet duidelijk worden aangegeven voor welk tijdsbestek, omstandigheden, doel en activiteiten het onderzoek geldt	0	Kan beter

De conclusie op basis van de algemeen wetenschappelijke eisen luidt dat het rapport nog wel beter kan, hoewel de basis goed is. Er kan nog duidelijker worden gemaakt hoe de dataverzameling heeft plaatsgevonden en hoe deze is geanalyseerd.

### Conclusie

Dit rapport voldoet geheel aan de eisen aan een auditingsrapport. Daarnaast is de basis gelegd om in de toekomst tot een evaluatieonderzoek te komen, wanneer enkele punten beter worden uitgewerkt. Wellicht is dit niet het doel van de onderzoekers of opdrachtgevers (de titel is ook beleidseffectrapportage), maar de data die in de loop der jaren zijn verzameld zijn te kostbaar om verloren te gaan. Het verdient dan ook de aanbeveling om door te gaan met de dataverzameling, om in een later stadium in elk geval vergelijkingsmateriaal te hebben (iets waar het in veel evaluatieonderzoek aan ontbreekt).

### 6.4 Interprovinciale Rapportage Milieu, Water, Landbouw en Natuur 2002

In de Interprovinciale Rapportage Milieu, Water, Landbouw en Natuur 2002 wordt opgesteld in opdracht van en door het Interprovinciaal Vakberaad Monitoring (IVM), die ook vertegenwoordigd is in het project groep. Naast het IVM zijn ook het RIVM en een drietal provincies vertegenwoordigt.

In het voorwoord in dit rapport wordt vermeld dat het gaat om de zesde opeenvolgende monitoringsrapportage van de gezamenlijke provincies en het Interprovinciaal Overleg (IPO). Het rapport wordt gezien als basis voor een beleidsevaluatie, die aanbevelingen levert voor bijsturing van het beleid.

Het rapport handelt over de ontwikkelingen rondom het thema milieu in verhouding tot verschillende subthema's zoals landelijk gebied, verkeer, energie en afval. Hierbij wordt per hoofdstuk de subthema's behandeld. Dit rapport is als geheel aan een meta-evaluatie onderwerpen, aangezien uitsplitsing naar hoofdstuk niks toevoegt aan de beoordeling van het rapport

Hoewel het rapport zelf als monitoringsonderzoek is ingestoken, zal blijken dat het verder reikt dan dat. Dit is opvallend, aangezien in veel rapporten dit omgekeerd is: de ambitie is groter dan de uiteindelijke uitvoering.

Opvallend is verder dat, hoewel het monitoringsrapport betreft, de begrippen beleid, monitoring en evaluatie en de onderlinge relaties in een min of meer theoretisch hoofdstuk worden toegelicht. Dit is begrijpelijk met het oog op de eerder genoemde functie van het

rapport om een basis te bieden voor beleidsevaluatie, maar niet noodzakelijk voor een monitoringsrapport.

De beoordeling op de lijst met criteria is hieronder beschreven.

### Monitoring

Criterion	Beoordeling	Opmerkingen
Het rapport geeft ontwikkelingen aan uit de beleidspraktijk door middel van metingen en observaties van relevante variabelen over een bepaald tijdsbestek en met een bepaald doel	+	Aan dit criterium wordt ruimschoots voldaan, de beoordeling hiervan is een +. Met andere woorden: er kan gesproken worden van een monitoringsrapportage.

Dit rapport geeft in de inleiding al aan dat het om een monitoringsrapport gaat, hetgeen terecht is aangezien aan deze voorwaarde met betrekking tot monitoring wordt voldaan.

### Auditing

Criterion	Beoordeling	Opmerkingen
Het rapport bevat één of meer duidelijk geformuleerde en beargumenteerde maatstaven en geeft een duidelijke beoordeling aan de hand van deze vermelde maatstaven	0	Hoewel het rapport slechts monitoring tot doel heeft, worden bij enkele thema's ook doelstellingen opgenomen
Het rapport geeft inzicht in de mate van doelbereiking	-	
Het rapport maakt ten aanzien van de effecten onderscheid tussen positief gewaardeerde effecten (baten of voordelen) en negatief gewaardeerde effecten (kosten of nadelen) en tracht deze met elkaar te vergelijken.	-	Er wordt wel onderscheid gemaakt tussen positieve en negatieve effecten, maar deze worden niet met elkaar vergeleken
Het rapport besteedt aandacht aan de neveneffecten van het beleid en geeft aan of deze negatief of positief worden gewaardeerd	-	

Dit rapport is inderdaad een monitoringsrapport, en geen auditing- of evaluatierapport. Het heeft weinig zin om de criteria die aan een evaluatierapport worden gesteld nog te doorlopen, aangezien aan de criteria over de auditing al grotendeels niet wordt voldaan.

### Legitimiteit

Het beleid moet inzicht geven in de mate waarin het beleid wordt gedragen door betrokken partijen.	0	Doordat in een projectgroep wordt gewerkt, wordt inzicht verkregen in het draagvlak, maar in het rapport zelf komt dit niet naar voren in harde doelstellingen
--	---	--

Tot slot wordt gekeken naar de algemene wetenschappelijke eisen.

### Algemeen wetenschappelijk

Criterion	Beoordeling	Opmerkingen
De wijze waarop de data zijn beoordeeld, geanalyseerd en geïnterpreteerd is duidelijk en expliciet omschreven in het rapport	+	
In het rapport moet aangegeven worden welke beperkingen het onderzoek kent, met betrekking tot de dataverzameling, methode, referentiemateriaal, etc. met het oog op de geldigheid van de conclusies	+	Erg goed
Het rapport moet gefocust zijn op de belangrijkste vragen (sleutelvragen) en de conclusies moeten in de eerste plaats hierop gericht zijn	+	Wordt aan voldaan, hoewel het rapport wel veel onderwerpen betreft, waardoor ook veel sleutelvragen aanwezig zijn
In de conclusie(s) van het rapport moet duidelijk worden aangegeven voor welk tijdsbestek, omstandigheden, doel en activiteiten het onderzoek geldt	+	

## Conclusie

Dit rapport is een monitoringsrapportage. In dit geval is het geen probleem dat het 'slechts' een monitoringsrapportage betreft, omdat het rapport ook niet beoogt om een auditing of evaluatierapport te zijn, hiervoor dient het eventueel als basis. Het rapport is wel een goed monitoringsrapport, het criterium voor monitoring wordt zeer goed uitgewerkt. Hierdoor is de verwachting dat het rapport inderdaad als input kan dienen voor een auditing- of evaluatierapport. Deze verwachting wordt versterkt doordat bij het opstellen van de monitoring al (gedeeltelijk) rekening is gehouden met doelstellingen in het beleid, hetgeen niet per se nodig is voor een monitoringsrapport. Het maakt het uitvoeren van een evaluatie echter wel eenvoudiger en beter.

Ook wanneer de algemeen wetenschappelijke eisen worden meegenomen, dan blijkt dit rapport daar niet minder dan positief op te scoren. Dit geeft aan dat deze 'Interprovinciale Rapportage Milieu, Water, Landbouw en Natuur 2002' een goed uitgevoerd monitoringsonderzoek is.

Wel moet vermeld worden dat het onderzoek grotendeels intern is uitgevoerd. Doordat opdrachtgever en uitvoerende instantie grotendeels dezelfde zijn, en ook de leden van de projectgroep binding hebben met de opdrachtgever, is de kritische vraag gerechtvaardigd in hoeverre het onderzoek objectief is uitgevoerd. Aangezien hier geen zicht op is, is het niet mogelijk om hier een oordeel over te geven.

## 6.5 Evaluatie van de kampeerregelgeving – 2002

Hoewel de kampeerregelgeving slechts voor een deel met ruimtelijke ordening te maken heeft, is dit rapport toch opgenomen in dit onderzoek. De reden hiervoor dat dit onderzoek is uitgevoerd door de organisatie "Research voor Beleid". Deze organisatie voert al geruime tijd beleidsevaluatieonderzoeken uit, waardoor de nieuwsgierigheid wordt gewekt of jarenlange ervaring zich uit in een betere kwaliteit van evaluatieonderzoek.

Het rapport dat hier is geëvalueerd heeft als titel "Evaluatie van de kampeerregelgeving" en is in 2002 opgesteld. In het rapport ligt de nadruk op het vaststellen van de doorwerking van de kampeerregelgeving in het gemeentelijk beleid. De nadruk op doorwerking is ook interessant wanneer gekeken wordt naar de opmerking van Arts dat juist doorwerking één van de belangrijkste onderwerpen van een evaluatieonderzoek zou moeten zijn. De vraag stond hierbij centraal of de Wet op de Openlucht recreatie hierbij voor meer beleidsvrijheid bij gemeenten en voor een grotere diversiteit van kampeervormen heeft gezorgd.

### Monitoring

Criterion	Beoordeling	Opmerkingen
Het rapport geeft ontwikkelingen aan uit de beleidspraktijk door middel van metingen en observaties van relevante variabelen over een bepaald tijdsbestek en met een bepaald doel	+	

### Auditing

Criterion	Beoordeling	Opmerkingen
Het rapport bevat één of meer duidelijk geformuleerde en beargumenteerde maatstaven en geeft een duidelijke beoordeling aan de hand van deze vermelde maatstaven	0	De doelstellingen zijn niet helder geformuleerd, meer beschrijvend dus niet geconcretiseerd. Dit bemoeilijkt de beoordeling. Het hele rapport is eigenlijk gefocust op draagvlak (legitimiteit) onder de betrokken partijen.
Het rapport geeft inzicht in de mate van doelbereiking	-	Lastig te meten
Het rapport maakt ten aanzien van de effecten onderscheid tussen positief gewaardeerde effecten (baten of voordelen) en negatief gewaardeerde effecten (kosten of nadelen) en tracht deze met elkaar te vergelijken.	0	Er wordt wel onderscheid gemaakt tussen positieve en negatieve effecten, maar deze worden niet met elkaar vergeleken

Het rapport besteedt aandacht aan de neveneffecten van het beleid en geeft aan of deze negatief of positief worden gewaardeerd	+	Dit rapport wijst ook op neveneffecten van het beleid.
--	---	--

Aangezien het rapport twee keer neutraal, één keer negatief en één keer positief scoort is de vraag gerechtvaardigd of hier gesproken kan worden over een auditingsrapport. Aangezien het doel van het rapport 'evaluatie' is dan mag verwacht worden dat het ook aan deze voorwaarden voldoet. Daarom worden, ondanks dat het rapport nauwelijks voldoende scoort op auditing, toch de evaluatiecriteria doorlopen. Zoals in de onderzoeksmethodiek al naar voren is gekomen, is de grens tussen de typen onderzoek, monitoring, auditing en evaluatie, niet altijd eenvoudig te trekken.

### Evaluatie

criterium	Beoordeling	Opmerkingen
Het rapport verwijst naar één of meer bepaalde beleidsdoelstellingen van bepaalde personen of organisaties en gaat na welke de uitwerking het beleid uiteindelijk heeft gehad of kan hebben	0	Hoewel de doeleinden vaag zijn, wordt goed omschreven welke invloed het beleid heeft
Het rapport geeft een verklaring voor het al dan niet bereiken van de beleidsdoelstellingen	0	Aangezien de doelstellingen niet gekwantificeerd zijn, kan het bereiken hiervan niet geheel bepaald worden
Het rapport geeft een beschrijving van het beleidsveld en de daarin optredende relaties	+	
Het rapport gaat na wat het beleid tot de oplossing van het probleem heeft bijgedragen of kan bijdragen (duidelijke probleemstelling ook nodig)	+	Doordat het beleidsveld goed is omschreven, is ook duidelijk hoe het beleid bijdraagt aan de verschillende 'problemen'
Het rapport verwijst naar één of meer bepaalde beleidsinstrumenten en gaat na in hoeverre deze hebben bijgedragen of kunnen bijdragen aan het bereiken van het doel waarvoor zij zijn ingezet	+	Er wordt duidelijk gemaakt welke regels en wetten er gelden en hoe verschillende instrumenten worden gebruikt
Het rapport moet inzicht geven in de interveniërende variabelen (externe invloeden), die van invloed kunnen zijn op het al dan niet bereiken van de beleidsdoelstellingen	+	
Het rapport geeft ten aanzien van de verklarende variabelen (beleidsuitvoering en bepaalde kenmerken van het beleidsveld) aan in hoeverre deze vatbaar zijn voor beïnvloeding door de beleidsinstantie, dus manipuleerbaar zijn	+	Er wordt aangegeven hoe de beleidsuitvoering en het beleidsveld kunnen worden beïnvloed en op welke manier dit het beste kan
Het rapport moet inzicht geven in de sterke en zwakke punten van het beleid	+	Deze worden expliciet genoemd en ook worden aanbevelingen gedaan hoe dit verbeteren
Het rapport moet aangeven in hoeverre de sterke punten van het beleid kunnen worden gebruikt om de zwakke punten te beperken of weg te werken	0	Aanbevelingen worden wel gegeven, maar niet zozeer hoe de sterke punten aangewend kunnen worden om de zwakke punten weg te werken
Het beleid moet inzicht geven in de mate waarin het beleid wordt gedragen door betrokken partijen.	+	

### Algemeen wetenschappelijk

criterium	Beoordeling	Opmerkingen
De wijze waarop de data zijn beoordeeld, geanalyseerd en geïnterpreteerd is duidelijk en expliciet omschreven in het rapport	+	
In het rapport moet aangegeven worden welke beperkingen het onderzoek kent, met betrekking tot de dataverzameling, methode, referentiemateriaal, etc. met het oog op de geldigheid van de conclusies	+	
Het rapport moet gefocust zijn op de belangrijkste vragen (sleutelvragen) en de conclusies moeten in de eerste plaats hierop gericht zijn	+	

In de conclusie(s) van het rapport moet duidelijk worden aangegeven voor welk tijdsbestek, omstandigheden, doel en activiteiten het onderzoek geldt	+	
---	---	--

### Conclusie

De algemene conclusie luidt dat dit een goed evaluatieonderzoek is, hoewel aan enkele eisen met betrekking tot auditing nog wel beter kan worden voldaan. Het is niet geheel duidelijk wat de precieze doelstellingen zijn van het onderzochte beleid. Hierdoor blijft de mate van doelbereiking onduidelijk. Dit is echter meer een tekortkoming van het beleid dan van het evaluatieonderzoek. Bij het opstellen van het beleid is te weinig rekening gehouden met de evalueerbaarheid ervan. Door de doelstellingen deze duidelijker te concretiseren en te kwantificeren kan hierop nog veel vooruitgang worden geboekt.

Hoewel rapport niet optimaal scoort op de auditing, is de uiteindelijke beoordeling van het evaluatieonderzoek toch positief. Dit komt doordat het bovengemiddeld scoort op de eisen die gesteld worden aan evaluatieonderzoek. Daarnaast valt op dat aan de algemene wetenschappelijke eisen wordt voldaan. Voor het feit dat niet geheel wordt voldaan aan de eisen met betrekking tot auditing ligt de oorzaak niet in het evaluatieonderzoek, maar in het beleid zelf (te vage doelstellingen).

De verklaring voor de betere kwaliteit van dit evaluatieonderzoek ligt waarschijnlijk in het feit dat het een onafhankelijk adviesbureau is, dat gespecialiseerd is in evaluatieonderzoek.

### 6.6 Evaluatieonderzoek wet- en regelgeving milieujaarverslagen

Dit evaluatieonderzoek is uitgevoerd door KPMG Sustainability, in samenwerking met onderzoeksinstituut NIPO. Dit schept hoge verwachtingen, aangezien dit beide gerenommeerde instanties zijn. Opdrachtgever is het Ministerie van VROM. Het onderzoek is begeleid door een begeleidingscommissie met daarin een vertegenwoordiging vanuit het ministerie van VROM, VNO-NCW, de Unie van Waterschappen, het ministerie van Verkeer en Waterstaat, het Interprovinciaal Overleg (IPO), de Stichting Natuur en Milieu en KPMG Sustainability.

Er worden drie onderdelen onderzocht, namelijk de effecten van de wettelijke regeling milieuverlaglegging, de uitvoering van de regeling en doelgroep verslagplichtige bedrijven (in hoeverre vallen de juiste bedrijven onder de regeling).

### Monitoring

criterium	Beoordeling	Opmerkingen
Het rapport geeft ontwikkelingen aan uit de beleidspraktijk door middel van metingen en observaties van relevante variabelen over een bepaald tijdsbestek en met een bepaald doel	+	

### Auditing

criterium	Beoordeling	Opmerkingen
Het rapport bevat één of meer duidelijk geformuleerde en beargumenteerde maatstaven en geeft een duidelijke beoordeling aan de hand van deze vermelde maatstaven	+	De doelstellingen zijn helder geformuleerd. Dit bemoeilijkt de beoordeling. Het hele rapport is eigenlijk gefocust op draagvlak (legitimiteit) onder de betrokken partijen.
Het rapport geeft inzicht in de mate van doelbereiking	+	
Het rapport maakt ten aanzien van de effecten onderscheid tussen positief gewaardeerde effecten (baten of voordelen) en negatief gewaardeerde effecten (kosten of nadelen) en tracht deze met elkaar te vergelijken.	0	Er wordt wel onderscheid gemaakt tussen positieve en negatieve effecten, maar deze worden niet met elkaar vergeleken
Het rapport besteedt aandacht aan de neveneffecten van het beleid en geeft aan of deze negatief of positief worden gewaardeerd	+	Goed uitgewerkt, expliciet genoemd

Dit onderzoek voldoet aan de eisen die worden gesteld aan een auditing. Wat opvalt is dat ook hier de vergelijking tussen positief en negatief gewaardeerde effecten ontbreekt. Daarnaast worden de neveneffecten niet als zodanig aangeduid, maar zijn ze wel opgenomen. Zoals “bij de geïnterviewde stakeholders wordt een vergroot inzicht in de bedrijfsvoering door het publieksverslag niet als zodanig herkend. Wel geven zij aan dat er door de verslagplicht meer milieugegevens beschikbaar komen en dat de milieuprestaties van een bedrijf daarmee inzichtelijker worden” (p. 70). Dus het hoofddoel (effect) is niet als zodanig gehaald, maar er worden wel andere effecten (neveneffecten) herkend.

### Evaluatie

criterium	Beoordeling	Opmerkingen
Het rapport verwijst naar één of meer bepaalde beleidsdoeleinden van bepaalde personen of organisaties en gaat na welke de uitwerking het beleid uiteindelijk heeft gehad of kan hebben	0	De doeleinden zijn goed omschreven, de invloed van het beleid kan duidelijker worden beschreven
Het rapport geeft een verklaring voor het al dan niet bereiken van de beleidsdoelstellingen	+	Het doet een poging tot verklaren aan de hand van enquêtes
Het rapport geeft een beschrijving van het beleidsveld en de daarin optredende relaties	-	
Het rapport gaat na wat het beleid tot de oplossing van het probleem heeft bijgedragen of kan bijdragen (duidelijke probleemstelling ook nodig)	0	Dit wordt niet geheel duidelijk
Het rapport verwijst naar één of meer bepaalde beleidsinstrumenten en gaat na in hoeverre deze hebben bijgedragen of kunnen bijdragen aan het bereiken van het doel waarvoor zij zijn ingezet	0	Er wordt duidelijk gemaakt welke regels en wetten. Onduidelijk blijft hoe de instrumenten precies worden ingezet
Het rapport moet inzicht geven in de interveniërende variabelen (externe invloeden), die van invloed kunnen zijn op het al dan niet bereiken van de beleidsdoelstellingen	-	
Het rapport geeft ten aanzien van de verklarende variabelen (beleidsuitvoering en bepaalde kenmerken van het beleidsveld) aan in hoeverre deze vatbaar zijn voor beïnvloeding door de beleidsinstantie, dus manipuleerbaar zijn	+	Er wordt aangegeven hoe de beleidsuitvoering kan worden beïnvloed en op welke manier dit het beste kan. Het beleidsveld is al eerder als negatief beoordeeld dus zou 'dubbel tellen' wanneer dit weer als negatief wordt aangemerkt
Het rapport moet inzicht geven in de sterke en zwakke punten van het beleid	0	Deze worden niet expliciet genoemd, hoewel wel het rapport wel verkapte aanbevelingen bevat hoe het beleid te verbeteren.
Het rapport moet aangeven in hoeverre de sterke punten van het beleid kunnen worden gebruikt om de zwakke punten te beperken of weg te werken	-	
Het beleid moet inzicht geven in de mate waarin het beleid wordt gedragen door betrokken partijen.	0	Hoewel wel wordt gekeken naar de ervaringen van betrokkenen met het beleid, zo dit onderdeel toch beter uitgewerkt kunnen worden.

Het evaluatieonderzoek is niet geheel volgens de regels uitgevoerd die hiervoor gelden. Vooral het inzicht welke rol het beleid heeft gespeeld en welke sterke en zwakke punten het heeft blijft onderbelicht. Daartegenover staat dat het rapport inzicht geeft in de neveneffecten van het beleid. Wanneer dieper in zou worden gegaan op het beleid en beleidsveld dan is zal de beoordeling van het rapport hoger uitvallen.

**Algemeen wetenschappelijk**

Criterion	Beoordeling	Opmerkingen
De wijze waarop de data zijn beoordeeld, geanalyseerd en geïnterpreteerd is duidelijk en expliciet omschreven in het rapport	0	Kan beter
In het rapport moet aangegeven worden welke beperkingen het onderzoek kent, met betrekking tot de dataverzameling, methode, referentiemateriaal, etc. met het oog op de geldigheid van de conclusies	0	Er worden wel beperkingen aangegeven, maar niet in verband met de conclusies. Daarom geen +
Het rapport moet gefocust zijn op de belangrijkste vragen (sleutelvragen) en de conclusies moeten in de eerste plaats hierop gericht zijn	+	Deze worden in de eerste pagina's van het rapport expliciet genoemd, en in de conclusies nog eens herhaald
In de conclusie(s) van het rapport moet duidelijk worden aangegeven voor welk tijdsbestek, omstandigheden, doel en activiteiten het onderzoek geldt	0	Hoewel de conclusies helder zijn, wordt niet aangegeven voor welke tijdsbestek, omstandigheden, etc. ze gelden

**Conclusie**

Dit 'evaluatieonderzoek' neigt meer naar een auditing dan naar daadwerkelijk evaluatie. Dit wordt met name veroorzaakt door gebrek aan inzicht in het beleidsveld. Daardoor worden niet altijd even duidelijke verklaringen gegeven voor de waargenomen effecten en de invloed van het beleid hierop. De materie is complex, maar dat telt niet mee in bij het bepalen van de kwaliteit van evaluatieonderzoek.

Op basis van de beoordeling op grond van algemeen wetenschappelijke eisen kan gesteld worden dat het rapport nog wel voor verbetering vatbaar is, vooral in de beschrijving van de omstandigheden waaronder de conclusies zijn getrokken en de data zijn verzameld en geanalyseerd. Dit blijft onderbelicht.

**6.7 Monitoring en evaluatie deregulering taxivervoer 1999-2001**

Dit rapport is, in opdracht van het Ministerie van Verkeer en Waterstaat, uitgevoerd door KPMG Bureau voor Economische Argumentatie in samenwerking met NIPO. Het rapport beschrijft de richting en omvang van geconstateerde ontwikkelingen in de taximarkt in de periode 1999-2001. Daarnaast probeert het de gevonden ontwikkelingen te verklaren om verslag te kunnen doen van de doeltreffendheid en de effecten in de praktijk van het nieuwe taxibeleid dat op 1 januari 2000 is ingevoerd.

Het rapport bestaat uit twee onderdelen, namelijk een monitor en een evaluatie. Het feit dat dit expliciet in de titel wordt genoemd is reden geweest om deze studie mee te nemen in dit onderzoek. De vraag is of in het rapport duidelijk de scheiding valt aan te geven tussen de monitor en de evaluatie. In de inleiding wordt hier alvast een voorschot op gegeven.

De monitor heeft een tweeledig doel. In de eerste plaats dient het de richting en omvang van de ontwikkelingen in de taximarkt te beschrijven. In de tweede plaats dient het de gevonden ontwikkelingen te verklaren teneinde periodiek verslag te kunnen doen van de doeltreffendheid en de effecten van de nieuwe wetgeving. Om een zo compleet mogelijk beeld van de ontwikkelingen in de taxibranche te krijgen worden jaarlijks een aantal deelonderzoeken uitgevoerd, te weten:

- het consumentenonderzoek, gericht op het in kaart brengen van ontwikkelingen aan de vraagzijde van de taximarkt in de jaren 1999-2001;
- het bedrijvenonderzoek, gericht op de ontwikkelingen aan de aanbodzijde in 1999-2001;
- het onderzoek naar de ontwikkelingen op het gebied van vergunningverlening, handhaving en tarieven in de jaren 1999-2001;
- het onderzoek naar de aanpassing van de infrastructuur.



De evaluatie van de effecten van het nieuw ingevoerde beleid wordt uitgevoerd aan de hand van 11 indicatoren. Ten behoeve van de monitor zijn 11 indicatoren gedefinieerd. Deze indicatoren, waarbij voor elke indicator is aangegeven welke ontwikkeling als gevolg van de deregulering wordt verwacht, zijn in het onderzoek uitgewerkt.

### Monitoring

Criterion	Beoordeling	Opmerkingen
Het rapport geeft ontwikkelingen aan uit de beleidspraktijk door middel van metingen en observaties van relevante variabelen over een bepaald tijdsbestek en met een bepaald doel	+	Aan dit criterium wordt voldaan, de beoordeling hiervan is een +. Met andere woorden: er kan gesproken worden van een monitoringsrapportage.

De verwachting dat het rapport minimaal aan het criterium met betrekking tot monitoring zou voldoen, komt uit. Het rapport spreekt al in de inleiding over criteria die afgeleid zijn uit de beleidspraktijk waaraan getoetst gaat worden.

### Auditing

Criterion	Beoordeling	Opmerkingen
Het rapport bevat één of meer duidelijk geformuleerde en beargumenteerde maatstaven en geeft een duidelijke beoordeling aan de hand van deze vermelde maatstaven	0	Er zijn wel doelstellingen uit het beleid overgenomen, maar deze blijven vaag. Dit komt deels doordat de doelstellingen moeilijk zijn te operationaliseren. Daarom beoordeling 0
Het rapport geeft inzicht in de mate van doelbereiking	+	Doordat de operationalisering niet optimaal is, is dit moeilijk realiseerbaar. Toch wordt een + gegeven, aangezien er concreet op de indicatoren wordt ingegaan
Het rapport maakt ten aanzien van de effecten onderscheid tussen positief gewaardeerde effecten (baten of voordelen) en negatief gewaardeerde effecten (kosten of nadelen) en tracht deze met elkaar te vergelijken.	0	Er wordt wel onderscheid gemaakt tussen positieve en negatieve effecten, maar deze worden niet met elkaar vergeleken
Het rapport besteedt aandacht aan de neveneffecten van het beleid en geeft aan of deze negatief of positief worden gewaardeerd	+	Goed uitgewerkt, expliciet genoemd, zoals een toename van de omzet door meer vraag, maar tegelijkertijd dempende werking van dalende prijzen door toename concurrentie

Het onderdeel auditing wordt overwegend positief beoordeeld. Dit komt wel overeen met de doelstelling van het rapport, aangezien daarin evaluatie als doel wordt aangeduid. Het noemen van de neveneffecten is hierbij een punt dat goed scoort, er wordt niet alleen gekeken naar de verwachte (positieve) effecten, maar er wordt ook melding gemaakt van neveneffecten, zoals een daling van de omzet door toegenomen efficiëntie.

### Evaluatie

Criterion	Beoordeling	Opmerkingen
Het rapport verwijst naar één of meer bepaalde beleidsdoeleinden van bepaalde personen of organisaties en gaat na welke de uitwerking het beleid uiteindelijk heeft gehad of kan hebben	+	De doelstellingen zijn niet optimaal geoperationaliseerd, maar zijn goed omschreven. De uiteindelijke invloed wordt wel redelijk duidelijk gemaakt, alleen blijft de vraag in hoeverre geldige uitspraken gedaan kunnen worden in verband met de korte periode waarop het onderzoek betrekking heeft.

Het rapport geeft een verklaring voor het al dan niet bereiken van de beleidsdoelstellingen	0	Slechts tot op een bepaalde hoogte
Het rapport geeft een beschrijving van het beleidsveld en de daarin optredende relaties	-	
Het rapport gaat na wat het beleid tot de oplossing van het probleem heeft bijgedragen of kan bijdragen (duidelijke probleemstelling ook nodig)	0	Dit wordt niet geheel duidelijk
Het rapport verwijst naar één of meer bepaalde beleidsinstrumenten en gaat na in hoeverre deze hebben bijgedragen of kunnen bijdragen aan het bereiken van het doel waarvoor zij zijn ingezet	+	Beleidsmaatregelen en – instrumenten worden stuk voor stuk beschreven (geen garantie dat alle maatregelen en instrumenten ook daadwerkelijk zijn meegenomen)
Het rapport moet inzicht geven in de interveniërende variabelen (externe invloeden), die van invloed kunnen zijn op het al dan niet bereiken van de beleidsdoelstellingen	+	
Het rapport geeft ten aanzien van de verklarende variabelen (beleidsuitvoering en bepaalde kenmerken van het beleidsveld) aan in hoeverre deze vatbaar zijn voor beïnvloeding door de beleidsinstantie, dus manipuleerbaar zijn	0	De verklarende variabelen worden wel genoemd, maar de invloed van het beleid hierop blijft onduidelijk
Het rapport moet inzicht geven in de sterke en zwakke punten van het beleid	-	Deze worden niet expliciet genoemd
Het rapport moet aangeven in hoeverre de sterke punten van het beleid kunnen worden gebruikt om de zwakke punten te beperken of weg te werken	-	Hangt samen met het voorgaande punt
Het beleid moet inzicht geven in de mate waarin het beleid wordt gedragen door betrokken partijen.	-	Slecht uitgewerkt, terwijl hier juist mogelijkheden lagen omdat de doelgroep duidelijk is.

Wat opvalt is dat ook hier geen inzicht wordt gegeven in het beleidsveld en de daarin voorkomende relaties. Hierdoor blijven de uitspraken vaag, aannames kunnen niet echt hard worden gemaakt, waardoor ook harde uitspraken ontbreken. In de toekomst zou dit een punt van verbetering moeten zijn, maar aangezien het onderzoek pas sinds enkele jaren wordt uitgevoerd, wordt hier het voordeel van de twijfel gegeven. Er wordt de aanbeveling gedaan om in de toekomst nogmaals onderzoek uit te voeren.

### Algemeen wetenschappelijk

Criterion	Beoordeling	Opmerkingen
De wijze waarop de data zijn beoordeeld, geanalyseerd en geïnterpreteerd is duidelijk en expliciet omschreven in het rapport	0	Kan beter
In het rapport moet aangegeven worden welke beperkingen het onderzoek kent, met betrekking tot de dataverzameling, methode, referentiemateriaal, etc. met het oog op de geldigheid van de conclusies	+	Positief aan dit onderzoek is dat wordt aangegeven dat de periode waarover de monitoring en evaluatie is uitgevoerd te kort is om goede uitspraken te doen
Het rapport moet gefocust zijn op de belangrijkste vragen (sleutelvragen) en de conclusies moeten in de eerste plaats hierop gericht zijn	+	Deze worden in de eerste pagina's van het rapport expliciet genoemd, en in de conclusies nog eens herhaald
In de conclusie(s) van het rapport moet duidelijk worden aangegeven voor welk tijdsbestek, omstandigheden, doel en activiteiten het onderzoek geldt	0	Hoewel de conclusies helder zijn, wordt niet aangegeven voor welke tijdsbestek, omstandigheden, etc. ze gelden

De algemeen wetenschappelijk eisen zijn redelijk goed uitgewerkt, een kritische kanttekening is echter op zijn plaats. De conclusies blijven namelijk vaag en blijven beperkt tot enkele punten van het gehele onderzoek. Daarnaast wordt niet de aanbeveling gedaan hoe op de conclusies moet worden ingespeeld bij het aanpassen van het vigerende beleid. Dit zou beter uitgewerkt kunnen worden.

## Conclusie

Dit rapport is meer een auditing- dan een evaluatierapport hoewel aan enkele punten met betrekking tot evaluatie wel wordt voldaan. De korte periode tussen invoering van het beleid en uitvoering van de evaluatie heeft er mede voor gezorgd dat er in het onderzoek geen harde uitspraken worden gedaan over het functioneren van het beleid. Dit hoeft geen probleem te zijn, mits dit bij de volgende evaluatie wel gebeurt. Hier moet in de volgende evaluatie voldoende aandacht aan worden geschonken.

Daarnaast is het opvallend dat de legitimiteit van het beleid niet in kaart is gebracht, aangezien juist hier de doelgroep duidelijk is: de taxibranche. Door middel van een enquête zou hier veel duidelijkheid over kunnen worden verkregen.

## 6.8 Evaluatie van de Derde Kamer

Dit onderzoek is door Research en Beleid uitgevoerd, in opdracht van de Nationale Commissie voor Internationale Samenwerking en Duurzame Ontwikkeling. Van februari tot november 2003 zijn 120 Nederlanders en 30 burgers uit het 'Zuiden' betrokken geweest bij De Derde Kamer. De Derde Kamer is een driejarig project (2003 –2006), waarbij burgers discussiëren over internationale samenwerking en duurzame ontwikkeling. De Derde Kamer beoogt de belangstelling van politiek, publiek en journalistiek voor deze onderwerpen te vergroten door publiciteit te zoeken en beleidsvoorstellen te schrijven over internationale samenwerking en duurzame ontwikkeling. Met deze voorstellen probeert De Derde Kamer Nederlandse politici en beleidsmakers te stimuleren een bijdrage te leveren aan duurzame ontwikkeling elders.

Aangetekend moet worden dat dit een tussenevaluatie betreft. Doel van deze tussentijdse evaluatie is het geven van inzicht in mogelijke verbeterpunten voor het project 'De Derde Kamer'. Dit betekent dat er aandacht is voor de gerealiseerde resultaten tot nu toe, maar de nadruk ligt op de mogelijkheden om de opzet en uitvoering van het project in de komende jaren te optimaliseren. Deze doelstelling is niet strijdig met de doelstelling die bij evaluatie in algemene zin geldt. De evaluatie richt zich op de vraag in hoeverre met de huidige aanpak het project 'De Derde Kamer' op de goede weg is om de gestelde doelen te bereiken. Hiervoor worden de opzet, uitvoering en de tot dusver behaalde resultaten van het project geanalyseerd en de mening van betrokkenen hierover in kaart gebracht.

## Monitoring

criterium	Beoordeling	Opmerkingen
Het rapport geeft ontwikkelingen aan uit de beleidspraktijk door middel van metingen en observaties van relevante variabelen over een bepaald tijdsbestek en met een bepaald doel	+	Aan dit criterium wordt voldaan, de beoordeling hiervan is een +. Met andere woorden: er kan gesproken worden van een monitoringsrapportage.

## Auditing

criterium	Beoordeling	Opmerkingen
Het rapport bevat één of meer duidelijk geformuleerde en beargumenteerde maatstaven en geeft een duidelijke beoordeling aan de hand van deze vermelde maatstaven	0	De doelstelling zijn vaag en worden niet geoperationaliseerd. Er wordt een bepaalde schaal gehanteerd maar onduidelijk blijft waar deze op gebaseerd is en wat bepalend is (bijvoorbeeld: wat is tevreden?)
Het rapport geeft inzicht in de mate van doelbereiking	-	Er wordt geen vergelijking gemaakt tussen doelstelling en doelbereiking
Het rapport maakt ten aanzien van de effecten onderscheid tussen positief gewaardeerde effecten (baten of voordelen) en negatief gewaardeerde effecten (kosten of nadelen) en tracht deze met elkaar te vergelijken.	-	Aangezien niet ingegaan wordt op de doelstellingen
Het rapport besteedt aandacht aan de neveneffecten van	-	Neveneffecten worden niet

het beleid en geeft aan of deze negatief of positief worden gewaardeerd		genoemd
---	--	---------

Aangezien dit onderzoek (nog) niet voldoet aan de eisen die gelden bij auditing, heeft het verder doorlopen van beslisboom (fig. 6.1), om te bepalen hoe het onderzoek voldoet aan de evaluatiecriteria, weinig zin. Wel wordt het criterium 'legitimiteit' nog beoordeeld:

### Legitimiteit

Het beleid moet inzicht geven in de mate waarin het beleid wordt gedragen door betrokken partijen.	+	Dit is zeer goed uitgewerkt, door enquêtes wordt duidelijk in hoeverre betrokkenen tevreden zijn over het beleid.
--	---	---

Tot slot wordt het onderzoeksrapport beoordeeld op de algemeen wetenschappelijke eisen:

### Algemeen wetenschappelijk

Criterion	Beoordeling	Opmerkingen
De wijze waarop de data zijn beoordeeld, geanalyseerd en geïnterpreteerd is duidelijk en expliciet omschreven in het rapport	+	
In het rapport moet aangegeven worden welke beperkingen het onderzoek kent, met betrekking tot de dataverzameling, methode, referentiemateriaal, etc. met het oog op de geldigheid van de conclusies	0	Hier wordt tot op een bepaalde hoogte op ingegaan
Het rapport moet gefocust zijn op de belangrijkste vragen (sleutelvragen) en de conclusies moeten in de eerste plaats hierop gericht zijn	0	Hier en daar wordt de aandacht wat afgeleid van de kernzaken
In de conclusie(s) van het rapport moet duidelijk worden aangegeven voor welk tijdsbestek, omstandigheden, doel en activiteiten het onderzoek geldt	+	

Op basis van de algemeen wetenschappelijke eisen geldt dat het rapport voldoende scoort, hoewel ook enkele punten, zoals zijn weergegeven in de tabel, voor verbetering vatbaar zijn.

### Conclusie

Het onderzoek spitst zich toe op kwantitatieve indicatoren, deze zijn moeilijk meetbaar. Daarnaast wordt aan enquêtes veel waarde gehecht terwijl niet duidelijk in hoeverre de geënquêteerden capabel geacht mogen worden om een goed oordeel te kunnen vellen. Ook de mate van subjectiviteit maakt enquêtes minder geschikt als methode om 'harde' uitspraken te doen. Ondanks deze 'negatieve' oordelen heeft het onderzoek wel waarde als monitoringsrapport en om in het vervolg als referentiemateriaal te dienen. Hieruit voortkomend is dan ook het advies om door te gaan met dit onderzoek. Het gevaar is dat in de komende jaren de groep van geënquêteerden van samenstelling verandert of dat de vraagstelling verandert. Daarnaast is de subjectieve opzet van het onderzoek een punt van aandacht. Bij een enquête is de 'waan van de dag' ook erg belangrijk. Hierdoor is het belangrijk om het tijdstip van de enquête ook te vermelden en aan te houden bij het vervolgonderzoek.

Voordeel van deze methode van onderzoek is het goede inzicht dat wordt verkregen in de mate van legitimiteit van het beleid.

Ondanks dat er in de inleiding van dit onderzoek al wordt gesproken van een tussenevaluatie, dus dat niet over alle indicatoren data voorhanden zijn en het dus moeilijk is om geldige uitspraken te doen, valt het oordeel negatief uit voor dit evaluatieonderzoek. Het is namelijk geen evaluatie, maar een monitoring. Het voldoet (nog) niet aan de eisen die worden gesteld aan auditing of evaluatie. De opzet van het onderzoek laat dit ook zien. Dit hoeft geen probleem te zijn, alleen de titel zou beter monitoring in plaats van evaluatie kunnen luiden.

## 6.9 Conclusie beoordeling evaluatieonderzoeken

Om dit onderzoek overzichtelijk te houden volgt hieronder in tabel 1 nogmaals de beoordeling van de evaluatieonderzoeken. De kolomtitels corresponderen met de volgorde van de evaluatieonderzoeken:

1. Beleids-effectrapportage Verkeer en Vervoer Noord-Nederland, periode 1986-2001;
2. Beleids-effectrapportage Verkeer en Vervoer 1999;
3. Interprovinciale Rapportage Milieu, Water, Landbouw en Natuur 2002;
4. Evaluatie van de kampeerregelgeving – 2002;
5. Evaluatieonderzoek wet- en regelgeving milieujaarverslagen;
6. Monitoring en evaluatie deregulering taxivervoer 1999-2001;
7. Evaluatie van de Derde Kamer.

### Monitoring

Criterion	1	2	3	4	5	6	7
Het rapport geeft ontwikkelingen aan uit de beleidspraktijk door middel van metingen en observaties van relevante variabelen over een bepaald tijdsbestek en met een bepaald doel	+	+	+	+	+	+	+

### Auditing

Criterion	1	2	3	4	5	6	7
Het rapport bevat één of meer duidelijk geformuleerde en beargumenteerde maatstaven en geeft een duidelijke beoordeling aan de hand van deze vermelde maatstaven	-	+	0	0	+	0	0
Het rapport geeft inzicht in de mate van doelbereiking	Nvt	+	-	-	+	+	-
Het rapport maakt ten aanzien van de effecten onderscheid tussen positief gewaardeerde effecten (baten of voordelen) en negatief gewaardeerde effecten (kosten of nadelen) en tracht deze met elkaar te vergelijken.	Nvt	0	-	0	0	0	-
Het rapport besteedt aandacht aan de neveneffecten van het beleid en geeft aan of deze negatief of positief worden gewaardeerd	Nvt	-	-	+	0	+	-

### Evaluatie

Criterion	1	2	3	4	5	6	7
Het rapport verwijst naar één of meer bepaalde beleidsdoeleinden van bepaalde personen of organisaties en gaat na welke de uitwerking het beleid uiteindelijk heeft gehad of kan hebben	Nvt	0	Nvt	0	0	+	Nvt
Het rapport geeft een verklaring voor het al dan niet bereiken van de beleidsdoelstellingen	Nvt	0	Nvt	0	+	0	Nvt
Het rapport geeft een beschrijving van het beleidsveld en de daarin optredende relaties	Nvt	-	Nvt	+	-	-	Nvt
Het rapport gaat na wat het beleid tot de oplossing van het probleem heeft bijgedragen of kan bijdragen (duidelijke probleemstelling ook nodig)	Nvt	0	Nvt	+	0	0	Nvt
Het rapport verwijst naar één of meer bepaalde beleidsinstrumenten en gaat na in hoeverre deze hebben bijgedragen of kunnen bijdragen aan het bereiken van het doel waarvoor zij zijn ingezet	Nvt	0	Nvt	+	0	+	Nvt
Het rapport moet inzicht geven in de interveniërende variabelen (externe invloeden), die van invloed kunnen zijn op het al dan niet bereiken van de beleidsdoelstellingen	Nvt	0	Nvt	+	-	+	Nvt
Het rapport geeft ten aanzien van de verklarende variabelen (beleidsuitvoering en bepaalde kenmerken van het beleidsveld) aan in hoeverre deze vatbaar zijn voor beïnvloeding door de beleidsinstantie, dus manipuleerbaar zijn	Nvt	-	Nvt	+	+	0	Nvt
Het rapport moet inzicht geven in de sterke en zwakke punten van het beleid	Nvt	0	Nvt	+	0	-	Nvt
Het rapport moet aangeven in hoeverre de sterke punten van het beleid kunnen worden gebruikt om de zwakke punten te beperken of weg te werken	Nvt	-	Nvt	0	-	-	Nvt

Het rapport moet inzicht geven in de mate waarin het beleid door wordt gedragen door betrokken partijen	-	0	0	+	0	-	+
---	---	---	---	---	---	---	---

### Algemeen wetenschappelijk

Criterion	1	2	3	4	5	6	7
De wijze waarop de data zijn beoordeeld, geanalyseerd en geïnterpreteerd is duidelijk en expliciet omschreven in het rapport	-	0	+	+	0	0	+
In het rapport moet aangegeven worden welke beperkingen het onderzoek kent, met betrekking tot de dataverzameling, methode, referentiemateriaal, etc. met het oog op de geldigheid van de conclusies	-	0	+	+	0	+	0
Het rapport moet gefocust zijn op de belangrijkste vragen (sleutelvragen) en de conclusies moeten in de eerste plaats hierop gericht zijn	-	+	+	+	+	+	0
In de conclusie(s) van het rapport moet duidelijk worden aangegeven voor welk tijdsbestek, omstandigheden, doel en activiteiten het onderzoek geldt	-	0	+	+	0	0	+

**Tabel 1** Overzicht beoordeling evaluatieonderzoeken

Bij het beoordelen van de in dit hoofdstuk genoemde evaluatieonderzoeken vallen enkele punten op. Dit zijn enerzijds punten die duidelijk voorkomen uit het evaluatieonderzoek zelf, anderzijds zijn dit punten die algemeen (bij meer evaluatieonderzoeken) voorkomen, waardoor hierover meer algemene uitspraken gedaan kunnen worden over de verbeterpunten van beleidsevaluatie. De conclusies op basis van tabel 1 luiden als volgt:

- Om te beginnen zijn alle geëvalueerde evaluatieonderzoeken minimaal allemaal een monitoring. Dat ligt in de lijn der verwachting, aangezien de onderzoeken allemaal een bepaald niveau van evaluatie in de titel bevatten en uit de inleidingen duidelijk blijkt dat op zijn minst een evaluatie wordt nagestreefd;
- Het valt op dat de onderzochte beleidsevaluatieonderzoeken slechts voor een deel voldoen aan de eisen die kunnen worden gesteld aan een evaluatieonderzoek. Dit valt te verklaren door de eerder genoemde valkuilen die zich laten gelden. Zo worden veel van de onderzoeken uitgevoerd in eigen beheer, wat merkwaardig genoemd mag worden met het oog op objectiviteit. Of de rapporten worden zijn wel door onafhankelijke bureaus uitgevoerd, maar (wellicht door het kostenaspect) ontbreken de verklaringen voor vastgestelde ontwikkelingen;
- Slechts één van de onderzochte evaluatieonderzoeken voldoet aan de eis dat inzicht moet worden gegeven in het beleidsveld en de daarbinnen optredende relaties.
- In enkele gevallen, zoals de beleidseffectrapportage verkeer en vervoer, is het beleid tijdens de evaluatie gewijzigd. Dit is bijzonder ergerlijk voor de evaluator, omdat verschillen tussen metingen, of juist het ontbreken daarvan, dan immers mogelijk ook toe te schrijven zijn aan de veranderingen in het beleid (Dekker, 2004). Hierdoor bestaat de kans dat de evaluator geen sluitend antwoord kan geven op de vraag van de opdrachtgever of het beleid nu wel of niet werkt of gewerkt heeft;
- Een fout die bij veel evaluatieonderzoek wordt gemaakt, het vaag omschrijven van de doelstellingen (Scheerens, 1983; Bressers, 1994; Ministerie van Financiën, 1993), doet zich bij enkele evaluatieonderzoeken voor, zoals bij de "Evaluatie van de kampeerregelgeving -2002" en de "Monitoring en evaluatie deregulering taxivervoer 1999-2001". Hierdoor is het bij voorbaat kansloos om een gedegen evaluatie uit te voeren. In het evaluatieonderzoek moet duidelijk zijn welke doelstellingen nagestreefd worden en waarop beoordeeld gaat worden;

- Bij het vaststellen van de effectiviteit wordt vaak niet aangegeven dat daadwerkelijk het beleid is verantwoordelijk is geweest een verandering. Wellicht lijkt op het eerste gezicht het beleid helemaal niet te werken, bijvoorbeeld een doelstelling is bij lange na niet gehaald, maar wie weet was zonder de invoering van het beleid het nog wel slechter gesteld met de het bereiken van de desbetreffende doelstelling. Deze bijdrage van beleid is misschien wel het moeilijkste vast te stellen. De constatering hangt voor een groot deel samen met het feit dat het beleidsveld slechts summier is uitgewerkt of zelfs ontbreekt. Doordat het inzicht in het beleidsveld ontbreekt, kan onmogelijk worden vastgesteld in hoeverre het beleid heeft bijgedragen aan de doelstellingen;
- Door Ziegelaar (2004) wordt een punt van kritiek aangegeven dat veel voorkomt bij evaluatieonderzoek en hier ook geldt, namelijk het feit dat veel onderzoek met een (quasi-) experimentele opzet is uitgevoerd. De experimentele methode is er op gericht de mate van het effect van een bepaalde maatregel vast te stellen. Dit gebeurt door de vergelijking van een experimentele groep, die onderworpen wordt aan de maatregel, met een controlegroep waarvoor dat niet geldt. Dit houdt in dat een evaluatie wordt uitgevoerd door de experimentele groep voor de invoering van het beleid te vergelijken met de groep na de invoering en looptijd van het beleid. Hierbij blijft het tussenliggende traject onderbelicht. De zwakte van deze opzet is dat het niet zichtbaar wordt wat er gebeurt tussen input en output. Er wordt slechts gekeken naar doelstelling en uitkomst, zonder dat ingegaan wordt op de oorzaken. Hierdoor wordt de onderliggende werking van een onderliggende maatregel niet blootgelegd. De waarde van dit type onderzoek is beperkt, omdat de 'zwarte doos' tussen input en output dicht blijft en daardoor onduidelijk is hoe het mechanisme werkt en van welke contextuele variabelen dat afhankelijk is. Deze kritiek hangt samen met het ontbreken van inzicht in het beleidsveld. Dit inzicht in het beleidsveld wordt in de meeste gevallen niet gegeven, waardoor de waarde van het evaluatieonderzoek beperkt blijft;
- Wanneer gekeken wordt naar de opdrachtnemer, dan is het opvallend dat de methodiek bij de onderzoeken die zijn uitgevoerd door Research voor Beleid goed in elkaar steekt, terwijl bijvoorbeeld het onderzoek dat door Grontmij Drenthe is uitgevoerd methodisch gezien niet juist is. Hoewel na één onderzoek de conclusie niet getrokken mag worden, lijkt het erop dat voor het uitvoeren van een evaluatieonderzoek specifieke kennis vereist is en deze bij gespecialiseerde bureaus aanwezig is. Het tijdsaspect zal in het geval van een commercieel adviesbureau als Grontmij ook een rol spelen bij de gebrekkige opzet van het evaluatieonderzoek;
- Tot slot is het opvallend dat de titel niet altijd de lading dekt. Het verschil tussen de titel en de daadwerkelijke uitvoering duidt erop dat voor de onderzoeker ook niet altijd duidelijk is wat de begrippen evaluatie, monitoring en effecten inhouden;

Deze punten gelden specifiek voor deze onderzoeken. In het slothoofdstuk wordt nader ingegaan op conclusie en aanbeveling die niet specifiek verband houden met de in dit hoofdstuk behandelde onderzoeken. Het is meer gericht op de conclusies zoals die ook op basis van het literatuuronderzoek zijn getrokken en de wijze waarmee hiermee kan worden omgegaan in evaluatieonderzoek.

## 7 Conclusie en aanbevelingen

### 7.1 Inleiding

De conclusies die in dit hoofdstuk zijn getrokken, zijn op basis van een combinatie van het uitgevoerde literatuuronderzoek en de geëvalueerde evaluatieonderzoeken. Hierdoor kan een antwoord worden gegeven op de onderzoeksvraag en de daarbij behorende doelstellingen. Voor de duidelijkheid wordt de onderzoeksvraag nog eens herhaald:

*Waar ligt de kwetsbaarheid van beleidsevaluatieonderzoek en hoe kan er een methode worden ontwikkeld om beleidsevaluatieonderzoek op een efficiënte manier te kunnen beoordelen en te verbeteren en kan hiermee ook het opstellen van beleid worden verbeterd?*

De daaruit voortvloeiende doelstellingen luiden als volgt:

1. Het in kaart brengen van de kwetsbaarheid van beleidsevaluatieonderzoek;
2. Het ontwikkelen van een methodiek voor het bepalen van de kwaliteit van beleidsevaluatieonderzoek;
3. Het toepasbaar maken van de methode voor het verbeteren van beleidsevaluatieonderzoek én beleid.

### 7.2 Conclusies

De conclusies op basis van dit voorliggende literatuur- en praktijkonderzoek zijn als volgt:

Een belangrijk onderdeel van evaluatie in algemene zin is het geven van een verklaring voor opgetreden effecten. Juist dit bepaalt het verschil tussen een evaluatie en een monitoring en auditing of evaluatie. Toch blijkt in de praktijk dit verklaren niet of onvoldoende te worden gedaan. Dit is een kwalijke zaak, aangezien het van groot belang is om inzicht te hebben in de omstandigheden waarin bepaalde effecten zich hebben voorgedaan. Zonder dit inzicht is het moeilijk om beleid accuraat bij te sturen. De oorzaak van ontbreken van een verklaring is moeilijk aan te geven, aangezien daar te weinig zicht op is, maar het is opvallend dat in veel evaluatieonderzoek geen inzicht wordt gegeven in het beleidsveld en de daarbinnen optredende relaties. Dit is een belangrijke factor voor het goed uitvoeren van evaluatieonderzoek. Vaak ontbreekt dit echter in het onderzoek, waardoor geen inzicht is de relatie tussen beleidsmaatregel en de optredende effecten. De fase tussen invoering van maatregelen en het al dan niet optreden van effecten heeft op deze manier veel weg van een 'black-box' (Ziegelaar, 2004): er is zicht op de input en output, maar de verklaring voor de output ontbreekt. Doordat dit achterwege wordt gelaten bestaat het gevaar dat betrokken partijen zich blind staren op de uitkomsten van het 'beleidsevaluatieonderzoek', terwijl dit in feite geen evaluatieonderzoek is, maar slechts een opsomming van waargenomen ontwikkelingen. Hiermee kan het beleid echter niet voldoende worden getoetst en beoordeeld.

Na bestudering van de onderzoekstechnische literatuur blijkt dat het onmogelijk is om een eenduidige methodiek te ontwikkelen die voor alle beleidsterreinen geldt. Wel is mogelijk gebleken om aan de hand van de checklist van Stufflebeam en criteria van onder andere Hoogerwerf en Arts een methodiek te ontwikkelen die tot een bepaalde hoogte de kwaliteit van een evaluatie bepaalt. Door hiermee rekening te houden bij het uitvoeren van een beleidsevaluatie kan in een vroeg stadium de hand worden geboden aan enkele valkuilen en problemen die zich voordoen bij beleidsevaluatieonderzoek.



De valkuilen en problemen zoals die zich voordoen bij evaluatieonderzoek zijn tijdens dit onderzoek ook naar voren gekomen. Zo is gebleken dat vooral de omschrijving en operationaliseren van doelstellingen in criteria vaak te wensen overlaat. Dit kan verschillende oorzaken hebben, maar een zich veel voordoende oorzaak is het vaag beschrijven van doelstellingen. Hierdoor zijn deze doelstellingen nauwelijks te evalueren, aangezien ze niet meetbaar kunnen worden gemaakt.

Door betrokkenheid bij het opstellen van de Beleidseffectrapportage Verkeer en Vervoer Noord-Nederland 1986-2001 is inzicht verkregen in de processen die spelen bij het opstellen van een beleidsevaluatiestudie. Een belangrijk punt is dat veel beleidsmakers nog steeds de illusie hebben dat zij ontwikkelingen die grotendeels afhankelijk zijn van marktwerking, kunnen sturen en beïnvloeden door beleid. Voor een deel zal dit nog lukken, maar op sommige punten zijn externe ontwikkelingen van grotere invloed dan het beleid, hoe confronterend dit ook mag zijn voor betrokken ambtenaren. Toch zal de ambtenaar bij het uitvoeren van een beleidsevaluatiestudie zoveel mogelijk terug willen zien dat zijn of haar beleid succesvol is. De druk die hierdoor aan de onderzoeker wordt opgelegd is dermate groot, dat van een onafhankelijk, objectief onderzoek nauwelijks sprake kan zijn. In het geval van de voornoemde beleidseffectrapportage ging het zelfs zo ver dat niet de onderzoekers bepaalden welke onderzoeksmethodiek gehanteerd werd en welke resultaten werden gerapporteerd, maar dat de opdrachtgever hiervoor de verantwoordelijkheid nam. Op deze manier zal het vrijwel onmogelijk blijken om een gedegen evaluatiestudie uit te voeren, die voldoet aan de voorwaarden van een 'goede' beleidsevaluatie. Dekker (2004) deelt deze mening. Hij geeft aan dat het van belang is om beleid en evaluatie vroegtijdig af te stemmen. Ondanks het feit dat in dit onderzoek deze constatering niet proefondervindelijk is vastgesteld, is dit een belangrijk aandachtspunt. Tegenwoordig moeten opdrachten boven een bepaald bedrag openbaar worden aanbesteed. Hierbij is direct contact tussen opdrachtgever en opdrachtnemer verboden. Dit staat een goede afstemming in de weg. Daarnaast speelt de tijdsdruk een grote rol in de beperkte afstemming: de tijd tussen het aanvragen en de 'deadline' voor uitbrengen van de offerte is vaak te kort. Het gevaar bestaat dat de onderzoeker in een spagaat belandt tussen geld en tijd wanneer voor hem niet duidelijk is wat de opdrachtgever precies verlangt.

Een ander punt is dat de onderzoeksvraag in de beginfase van de evaluatiestudie niet helder is. Hierdoor wordt lukraak informatie verzameld, zonder dat het uiteindelijke doel duidelijk is. Dit herbergt het gevaar dat er informatie wordt verzameld die er niet toe doet en afgeweken wordt van de kernthema's van het onderzoek. Het evaluatieonderzoek is dan meer een monitoringsrapport, zonder heldere doelstellingen, waarbij het uiteindelijk niet duidelijk wordt waartoe het onderzoek heeft gediend en zou moeten dienen.

Daarnaast valt het op dat het erop lijkt dat de kwaliteit van de geëvalueerde evaluatieonderzoeken afhankelijk is van de evaluatie uitvoerende instantie. De steekproef is echter te klein om hier harde conclusies aan te mogen verbinden. Aandachtspunt bij het beoordelen van evaluatieonderzoek is het onderscheid tussen intern en extern uitgevoerd beleidsevaluatieonderzoek. Er zijn twee dingen te zeggen over intern beleidsevaluatieonderzoek. Enerzijds kan een voordeel opleveren aangezien de persoon of instantie die de evaluatie uitvoert goed op de hoogte is van het beleid en alle relaties en processen die hierbij komen kijken. Anderzijds bestaat het gevaar dat de interne persoon of instantie bevooroordeeld en niet objectief te werk zal gaan. Een bepaalde mate van subjectiviteit is bij (wetenschappelijk) beleidsevaluatieonderzoek echter nooit helemaal uit te sluiten.

Hoewel in dit rapport een poging is gedaan om te komen tot een objectieve meting van evaluatieonderzoek, vragen sommige onderzoekers zich af of dit überhaupt mogelijk is

(Dekker, 2004). Zij zijn van mening dat een objectieve meting van effectiviteit van beleid niet mogelijk is. De consequentie hiervan is dat een objectieve meta-evaluatie, welke in dit onderzoek als leidraad heeft gediend, ook niet mogelijk is. Dekker (2004) is zeer kritisch naar deze zienswijze en manier van indekken door onderzoekers. Deze zogenaamde constructivisten zijn volgens hem vooral in de advieswereld te vinden. Hij is van mening dat het onderzoekstype dat door deze onderzoekers gehanteerd wordt, een subjectieve benadering gericht op overleg en ‘Habermasiaans’ consensus, niet per definitie tot een betere interpretatie leidt of zelfs maar iets toevoegt aan andere constructies. Dekker spreekt van ‘beleidssolipsisme<sup>18</sup>’: “ik heb het zo geconstrueerd, dus het zal wel goed wezen”. Hierbij is niet van belang wat er gebeurt, maar dat er wat gebeurt.

Ziegelaar (2004) merkt op dat bij de ontwikkeling van beleid meer rekening moet worden gehouden met de evalueerbaarheid ervan en dat te weinig naar de ‘counterfactual’ gezocht: de situatie zoals die geweest zou zijn als de beleidsmaatregel niet genomen was. In de evaluatieonderzoeken die in dit onderzoek zijn beoordeeld blijkt dat ook meer aandacht te behoeven. Er wordt nauwelijks gewezen op de situatie zoals die zou zijn geweest zonder invoering van de beleidsmaatregel. Beleidsmakers en beleidsonderzoekers hebben de gezamenlijke verantwoordelijkheid om de evalueerbaarheid van maatregelen te bevorderen. Daarnaast verdient het construeren van de ‘counterfactual’ meer aandacht, waarvoor vaak meer mogelijkheden zijn dan benut worden.

In dit rapport een soort checklist gegeven om een beleidsevaluatieonderzoek te evalueren. Uit literatuuronderzoek is echter gebleken dat het opstellen van zo’n checklist geen sinecure is. Stufflebeam is al enkele decennia bezig om een ‘checklist’ te ontwikkelen, maar door steeds veranderende inzichten en opvattingen is deze ook nooit ‘af’. De lijst met criteria kan dan ook gezien worden als een middel om evaluatieonderzoek te vereenvoudigen. Een groot voordeel van het gebruiken van dergelijke lijsten is dat al in een vroeg stadium duidelijk is waaraan de evaluatie moet voldoen en welke informatie hiervoor nodig. Met name in het geval van een onderzoek dat uitgevoerd wordt door een onafhankelijk, extern persoon of instantie kan dit in de relatie opdrachtgever – opdrachtnemer duidelijkheid verschaffen over de verantwoordelijkheden die een ieder heeft. Dit kan de efficiëntie van het onderzoek ten goede komen.

### 7.3 Aanbevelingen

Op basis van de conclusies uit paragraaf 7.1. kunnen aanbevelingen worden gedaan om het beleid, beleidsevaluatie en metaevaluatie te verbeteren. Deze aanbevelingen luiden als volgt: Het beleidsveld moet altijd uitgewerkt worden bij een evaluatieonderzoek. Dit is het namelijk het belangrijkste onderdeel bij het uitvoeren van een beleidsevaluatie. In de praktijk blijkt echter dat juist dit niet goed uitgewerkt is. De oorzaak hiervan is niet duidelijk, maar daar zou wellicht onderzoek naar kunnen worden gedaan. De aanbeveling luidt om bij elk evaluatieonderzoek zo goed mogelijk het beleidsveld in kaart te brengen. Hierdoor kunnen relaties in een vroeg stadium worden aangegeven en kunnen verklaringen makkelijker worden afgeleid. De moeite die dit in eerste kost, wordt later ruimschoots terugverdiend. In de eerste plaats door een kwalitatief beter onderzoek en in de tweede plaats door een toename van de efficiëntie bij het verdere verloop van het onderzoek. Het uitwerken van het beleidsveld is cruciaal voor het verklaren van al dan niet optredende effecten. Hierdoor kan er meer zicht komen op het optreden van neveneffecten.

Er moet meer aandacht komen voor de wisselwerking tussen beleid en evaluatie. Bij het opstellen van beleid moet meer rekening gehouden worden met de evalueerbaarheid ervan in een later stadium. Daarom moeten er concrete doelstellingen worden opgenomen, waarop het

---

<sup>18</sup> Solipsisme: filosofische leer die stelt dat alleen ons eigen ik en zijn bewustzijnsdaden bestaat

beleid later kan worden afgerekend. Vage doelstellingen, zoals die veel worden geformuleerd, moeten worden uitgesloten. Het beleid kan zo op een efficiëntere manier aan kwaliteit winnen, aangezien sneller duidelijk is welke doelstellingen meer aandacht nodig hebben of juist niet.

De onderzoeker en opdrachtgever moeten beter bekend zijn met het verschil tussen monitoring, auditing en evaluatie, en het onderscheid tussen de begrippen doelbereiking, doeltreffendheid en doelmatigheid. Door dit in een vroeg stadium duidelijk te maken, wordt voorkomen dat verwarring ontstaat over het onderzoek dat wordt uitgevoerd en het doel daarvan.

Aangezien het uitvoeren van evaluatieonderzoek complex is, is het aan te raden om het onderzoek uit te laten voeren door een gespecialiseerde instantie of onderzoeksbureau. Van belang hierbij is dat de evaluator toegang moet hebben tot alle relevante informatie.

Afstemming tussen evaluator en opdrachtgever over het type onderzoek en de daarbij te raadplegen data helpt om het onderzoek efficiënter te laten verlopen. Daarnaast kan door ook andere belanghebbende partijen in te schakelen voorkomen worden dat het onderzoek teveel gericht wordt op de doelstellingen die door de evaluator en opdrachtgever als belangrijk worden beschouwd en voorbijgaat aan andere relevante data.

Bij het uitvoeren van een evaluatieonderzoek is het aan te raden om de lijst met criteria waaraan een evaluatieonderzoek moet voldoen (meta-evaluatie) te raadplegen. Zo kan het evaluatieonderzoek vorm worden gegeven, zonder dat relevante criteria ontbreken.

Het zoeken naar de ‘counterfactual’ van beleid, oftewel wat zou de situatie zijn als het beleid niet was ingevoerd, verdient meer aandacht in evaluatieonderzoek. In dit onderzoek kan het niet meer worden toegevoegd, maar een extra criterium in figuur 5.2 zou kunnen luiden: “het rapport moet inzicht geven in de ‘counterfactual’ van beleid”

#### **7.4 Tot slot**

De conclusie op basis van de doelstellingen van dit rapport luidt dat er tot op een bepaalde wel een methode voor het evalueren van evaluatieonderzoek is te ontwikkelen, maar dat deze nooit compleet en af is. Hierbij is het duidelijk dat er algemene eisen zijn en eisen die meer specifiek zijn en bij elk onderzoek anders zijn. In de in onderzoek opgestelde methode kan wel op een snelle manier inzicht worden gekregen in het type onderzoek wat is uitgevoerd. Dat is al een eerste stap. De vraag gerechtvaardigd of het überhaupt mogelijk is om methodologisch gezien, 100% juist evaluatieonderzoek uit te voeren. Antwoord luidt waarschijnlijk nee, er kan nooit alle relevante informatie boven water gehaald worden. Dit hoeft geen probleem te zijn, maar het in de onderzoeken zou dit expliciet genoemd moeten worden. Sommige onderzoeken pretenderen de wijsheid in pacht te hebben terwijl hiervan nooit geheel sprake van kan zijn. De meerwaarde van evaluatieonderzoek ligt waarschijnlijk ook meer in het aanzwengelen van een (maatschappelijke) debat wanneer onvolkomenheden aan het licht komen, zodat gezamenlijk naar oplossing of verbetering gezocht kan worden, dan het voor 100% uitsluitel geven over beleidseffecten. Voor sommige wetenschappers of onderzoekers zal deze conclusie moeilijk verteerbaar zijn. Deze conclusie zal daarom ook wel weer onderwerp van discussie worden. Maar is dat niet waar het volgens dit onderzoek juist om draait: het aanzwengelen van een discussie, zodat uiteindelijk door verschillende partijen een oplossing of verbetering gevonden kan worden! Dit is overigens door de jaren heen ook niet anders geweest, na het bestuderen van de literatuur blijken er zoveel verschillende opvattingen ten aanzien van beleidsevaluatie de ronde te doen, dan het vrijwel onmogelijk is door de bomen het bos nog te zien. De kritiek die er in de literatuur terug te vinden is op de onderzoeksmethoden die in het evaluatieonderzoek worden gebruikt, maakt je bijna moedeloos. Tijdens dit onderzoek is steeds tegen wisselende inzichten aangelopen. Hierdoor is

de verwarring ook navenant toegenomen, want: welke methode is nu de beste? De conclusie van dit onderzoek luidt dan ook dat het onmogelijk is om één methode te ontwikkelen die voor elk situatie geschikt is. Dit sluit aan wat Perrin (2001) ook al constateerde: de beste lessen of methoden bestaan niet! Het lijkt nu of we er ons met een leien dakje vanaf maken, maar dat is niet het geval. Juist doordat tot deze conclusie wordt gekomen, maakt duidelijk dat er niet één werkwijze gevolgd is. De in dit onderzoek ontwikkelde methode is dan gebaseerd op verschillende werk- en zienswijzen. De kritiek hierop kan zijn dat voor de gulden middenweg is gekozen, en misschien is dat wel zo. Maar zoals Ziegelaar (2004) al vaststelde, er zijn verschillende onderzoeken naar dezelfde beleidsmaatregelen die tegenstrijdige uitkomsten geven. Volgens het ene onderzoek heeft de maatregel wel en volgens de andere geen effect. En welke is nu de juiste? Pawson (2004) geeft in een interview aan dat de nadruk niet zou moeten liggen op 'wat werkt' maar op 'wat werkt voor wie in welke omstandigheden'. Zo is ook de in dit onderzoek gehanteerde methode niet 'de beste methode'. Dit is na maanden onderzoek een op het eerste gezicht teleurstellende conclusie. Aan de andere kant is tijdens dit onderzoek wel duidelijk geworden dat de wisselwerking tussen beleid en beleidsevaluatie belangrijk is. Bij het opstellen van beleid zou rekening gehouden moeten worden met de evalueerbaarheid ervan. Beleid op zich wordt door dit onderzoek niet verbeterd, maar het als bewustzijn van beleidsmakers en –onderzoekers dat door bij opstellen van beleid rekening te houden met de evalueerbaarheid ervan de kwaliteit van zowel het beleid als de evaluatie toeneemt, dan is al veel gewonnen.

Dit rapport zal ook weer onderwerp van evaluatie zijn. De beoordeling op basis van de criteria die voor een scriptie gelden is in beginsel ook niet anders dan het doorlopen van een bepaalde checklist. Hieruit blijkt wel dat evaluatie veel vaker, min of meer onopgemerkt, wordt gedaan. Evaluatie is alledaagse bezigheid, of we daar nu bewust mee bezighouden of niet en niet alleen op het gebied van beleidsevaluatie.



## Literatuurlijst

- Abma, T. en R. in 't veld (red) (2001), *Handboek Beleidswetenschap*, Uitgeverij Boom, Amsterdam
- Adviesdienst Verkeer en Vervoer (1997), *Vervoer-Economische Verkenningen 1997-2002*, Sdu Uitgevers, Den Haag
- Arts, J. (1998), *EIA Follow-up: on the role of ex post evaluation in environmental impact assessment*, Geo Press, Groningen
- Baarda, D.B. en M.P.M de Goede (2001), *Basisboek Methoden en Technieken: Handleiding voor het opzetten en uitvoeren van onderzoek*, Wolters-Noordhoff bv, Groningen
- Beckman, K. (2004), *Kabinetsbeleid duurzame energie faalt*, in: *Het Financiële Dagblad* 19 juni 2004
- Blommestein, H.J., J.Th.A. Bressers, A. Hoogerwerf (red.) (1984), *Handboek beleidsevaluatie*, Samson Uitgeverij, Alphen aan den Rijn
- Botman, H. (2004), *Alarm om verdroging van natuur*, in: *Het Algemeen Dagblad* 22 mei 2004, PCM Interactive Media b.v.
- Bressers, J.Th.A en A. Hoogerwerf (1991), *Beleidsvaluatie: serie maatschappijbeelden*, Samson H.D. Tjeenk Willink, Alphen a/d Rijn
- Bressers, J.Th.A. (1983), *Beleidsdoelmatigheid en waterkwaliteitsbeleid: een bestuurskundig onderzoek*, Sneldruk Boulevard, Enschede
- Bressers, J.Th.A (2003), *Evaluatie van beleid*, in: *Overheidsbeleid: een inleiding in de beleidswetenschap*, Kluwer, Alphen a/d Rijn
- Bruinsma, F., J. van Dijk, C. Gorter (red.)(2002), *Mobiliteit en beleid*, Koninklijke Van Gorcum, Assen
- Chen, H.T., (1996), *A comprehensive typology for program evaluation*, in: *Evaluation Practice, Spring/Summer, 1996, vol. 17*
- Dekker, B. (2004), *Paddestoelen en valkuilen; faalfactoren in evaluatieonderzoek*, in: *Basis, themanummer Beleidsvaluatie, mei 2004*, Research voor Beleid, Leiden
- Doughlah, M, (1998), *Developing a concept of extension program evaluation*, in *Program and Development Evaluation*, University of Wisconsin, Wisconsin
- Feller, I. (2002), *Performance Measurement Redux*, in: *American Journal of Evaluation, Vol.23, No.4, p. 435-452*, American Evaluation Association, Elsevier Science Inc.
- Glasson, J., R. Therivel, A. Chadwick (1999), *Introduction to Environmental Impact Assessment: Principles and procedures, process, practice and prospects*, UCL Press Limited, London

Het Financiële Dagblad (2002), Overheid faalt bij reïntegratie van WAO'ers AMSTERDAM, 14 september 2002

Herweijer, M. en A. Hoogerwerf (1998) (red), Overheidsbeleid: Een inleiding in de beleidswetenschap, Samson Uitgeverij, Alphen aan den Rijn

Hoesel, P. van (2003), Beleidsonderzoek als professie: een lang gekoesterd ideaal, EIM, Zoetermeer

Honigh, M. (1985), Doeltreffend beleid: een empirische vergelijking van beleidssectoren, van Gorcum, Assen

Hoogerwerf, A en Zoutendijk (1991), in: *Beleidsvaluatie: serie maatschappijbeelden*, Samson H.D. Tjeenk Willink, Alphen a/d Rijn

Interprovinciaal Overleg (2002), Interprovinciale Rapportage Milieu, Water, Landbouw en Natuur 2002, Wilco bv, Amersfoort

Kickert, W.J.M. (1986), Overheidsplanning: Theorieën, technieken en beperkingen, Van Gorcum, Assen/Maastricht

Majchrzak, A. (1984), *Methods for policy research*, Sage Publications Inc., Beverly Hills, California USA

Mierlo, J.G.A. (2002), Beleidsfalen en beleidsevaluatie in de publieke sector, [www.fdewb.unimaas.nl/algec/staff/framepages/mierlo/VWEC-referaat.10.01.pdf](http://www.fdewb.unimaas.nl/algec/staff/framepages/mierlo/VWEC-referaat.10.01.pdf)

Ministerie van Financiën (1993), Beleidsonderzoek, het ontwikkelen en beoordelen van beleidsmaatregelen en -projecten, Sdu Uitgeverij, Den Haag

Ministerie van Financiën (1992), Evaluatiemethoden, een introductie, Sdu Uitgeverij, Den Haag

Ministerie van Financiën (1998), Handreiking Evaluatieonderzoek Ex Post, Sdu Uitgeverij, Den Haag

Ministerie van Financiën (2003), Handreiking Evaluatieonderzoek Ex Post, Sdu Uitgeverij, Den Haag

Ministerie van Financiën (1999), Heeft beleid effect?- Enkele benaderingen voor de evaluatie van bestaand beleid, Sdu Uitgeverij, Den Haag

Ministerie van Verkeer en Waterstaat (2000), Beleidseffectrapportage 1999: Beleidseffectmeting Verkeer en Vervoer, Van de Ridder bv, Nijkerk

Ministerie van Verkeer en Waterstaat (2002), Monitoring en evaluatie deregulering taxivervoer 1999-2001, Den Haag

Ministerie van VROM (2003), Evaluatieonderzoek wet- en regelgeving milieujaarverslagen, Den Haag

Morris, P. en R. Therivel (Ed.) (2001), *Methods of Environmental Impact Assessment*, Spon Press, London, England

- Niekerk, F. (2000), Het effect gerapporteerd: De gebruikswaarde van effectrapportages voor de planning van verkeersinfrastructuur, Geo Pers, Groningen
- NRC Handelsblad (2003), Kamer heeft genoeg van 'blikken vol beleid', 19 juni 2003, PCM Interactive Media b.v.
- NRC Handelsblad (2003), Tekort aan ouderen-woningen, 6 juni 2003, PCM Interactive Media b.v.
- Palumbo, D.J. (Ed.)(1987), The Politics of Program Evaluation, Sage Yearbooks in Politics and Public Policy, Sage Publications, Beverly Hills, California, USA
- Patton, M.Q. (2001), Evaluation, Knowledge Management, Best Practises and High Quality Lessons Learned, in: *American Journal of Evaluation*, Vol. 22, No.3, 2001
- Pawson, R. (2004), Pawson's pleidooi voor realistische beleidsevaluaties, in: *Basis, themanummer Beleidsevaluatie, mei 2004*, Research voor Beleid, Leiden
- Perrin, B. (2001), Commentary: Making Yourself – and Evaluation – Useful, in: *American Journal of Evaluation*, vol. 22, nr. 2, 2001
- Research voor Beleid (2003), Evaluatie van de Derde Kamer, Leiden
- Research voor Beleid (2002), Evaluatie van kampeerregelgeving, Leiden
- Rogers, S. (1999), Performance Management in Local Government, Financial Times Professional Limited, London
- Rosenthal, U et al. (1996), Openbaar bestuur: Beleid, organisatie en politiek, Samson H.D. Tjeenk Willink, Alphen aan den Rijn
- Samenwerkingsverband Noord-Nederland (2002), Beleidseffectrapportage Verkeer & Vervoer Noord-Nederland, periode 1986-2001: een overzicht van de effecten van het beleid in Noord-Nederland, Grontmij Drenthe, Assen
- Scheerens, J. (1983), Evaluatie-onderzoek en beleid: methodologische en organisatorische aspecten, Stichting voor Onderzoek van het Onderwijs, Flevodruk, Harlingen
- Schendelen, R. van (2002), Roosteren van de overheid helpt niet, in: De Volkskrant 28 december 2002, PCM Interactive Media b.v.
- Scriven, M. (2000a), Key Evaluation Checklist, <http://www.wmich.edu/evalctr/checklists>
- Scriven, M. (2000b), The logic and methodology of checklists, Claremont Graduate University
- Shadish, W. (1998), Some Evaluation Questions in *ERIC/AE Digest Series*, Catholic University of America, Washington
- Shepard, L.A. (1977), A checklist for evaluating large-scale assessment programs, Paper #9, Occasional Paper Series, The Evaluation Center, Western Michigan University
- Stokking, K.M. (1984), Interpreteren en evalueren: methodologie rondom de uitkomsten van onderzoek, Van Loghum Slaterus, Deventer



Stufflebeam, D.L. (2001), Evaluation Checklist: Practical Tools for Guiding and Judging Evaluations, in: *American Journal of Evaluation*, Vol.22, No.1, 2001, p.71-79, American Evaluation Association, Elsevier Science Inc.

Stufflebeam, D.L. (1999), Evaluation plans and operations checklist, <http://www.wmich.edu/evalctr/checklists>

Stufflebeam, D.L. (1974), Meta-evaluation

Stufflebeam, D.L. (1999), Program evaluations metaevaluation checklist, [http://www.wmich.edu/evalctr/checklists/program\\_metaeval\\_10point.pdf](http://www.wmich.edu/evalctr/checklists/program_metaeval_10point.pdf)

Stufflebeam, D.L. (2001), The Metaevaluation Imperative, in: *American Journal of Evaluation*, vol. 22, nr. 2, 2001

Trommelen, J. (2004), De aal wordt onbetaalbaar, in: De Volkskrant 24 maart 2004, PCM Interactive Media b.v.

Vartiainen, P., Evaluation methods and comparative study, University of Vaarsa, Vaarsa.

Voogd, H. (1995a), Facetten van de planologie, Samson H.D. Tjeenk Willink, Alphen aan den Rijn

Voogd, H. (1995b), Methodologie van ruimtelijke planning, Samson H.D. Tjeenk Willink, Alphen aan de Rijn

Vries, de, G.J. (2000), Beleidsdynamica als sociale constructie: een onderzoek naar doorwerking van beleidsevaluatie en beleidsadvisering, Eburon, Delft

Weiss, C.H. (2002), Patton's (2001) Lessons Learned: A Comment, in: *American Journal of Evaluation* Vol. 23, No.2, 2002

Walberg, H. J., Lu, D., Niemiec, R. P., & Walberg, H. J., III. (1997). A technique for objective analysis of qualitative evaluation data. *Evaluation Practice*, 18, 55-64

Wit, de J. en H. van Gent (2001), Economie en transport, Uitgeverij Lemma, Utrecht

Ziegelaar, A. (2004), Licht in de zwarte doos; naar een evenwichtige methodologie van beleidsevaluatie, in: *Basis, themanummer Beleidsevaluatie, mei 2004*, Research voor Beleid, Leiden

## **BIJLAGEN**