

EEN EXTERNE BEOORDELING VAN IN 1998 GEPUBLICEERD WODC-ONDERZOEK: HET AUDITRAPPORT

Michiel Herweijer, Peter van der Heijden, Jan Leijten, Ineke Haen.

Inleiding

Om de Adviesraad van het WODC in staat te stellen zich een oordeel te vormen over de kwaliteit van het onderzoek dat onder de verantwoordelijkheid van het WODC naar buiten wordt gebracht, is door de directeur van het WODC, prof.dr.H.G. van de Bunt, op 12 maart 1999 een onafhankelijke visitatiecommissie ingesteld. Het was de bedoeling dat deze externe commissie haar oordeel over het WODC-onderzoek en de eventuele suggesties tot verbetering voor 1 september 1999 zou voorleggen aan de Adviesraad van het WODC.

De commissie van externe beoordelaars werd gevormd door vier leden: prof.dr.I. Haen (hoogleraar criminologie, University of Nebraska), prof.dr.P.G.M. van der Heijden, hoogleraar statistiek ten behoeve van de sociale wetenschappen, Universiteit van Utrecht), prof.dr.M. Herweijer (hoogleraar bestuurskunde, Rijksuniversiteit Groningen) en prof.mr.J. Leijten (emeritus hoogleraar civiel procesrecht Nijmegen en Advocaat-Generaal bij de Hoge Raad). De commissie is in de periode mei-augustus 1999 drie keer bijeen geweest.

Om tot een kwaliteitsoordeel te komen heeft de commissie een kleine greep gedaan uit het totale aanbod van het in 1998 afgeronde WODC-onderzoek. Bij de selectie van nader te beoordelen onderzoeken is een onderscheid gemaakt tussen 'interne onderzoeken' en 'externe onderzoeken'. Ten aanzien van deze selectie heeft prof.dr.H.G. van de Bunt geadviseerd.

Uit het Jaarverslag 1998 van het WODC kwam naar voren dat er in dit verslagjaar 1998 15 *interne onderzoeken* werden afgerond. Acht daarvan (de meer omvangrijke onderzoeken) werden gepubliceerd in de reeks 'Onderzoek en Beleid'. Zeven daarvan (de minder omvangrijke) werden gepubliceerd in de reeks 'Onderzoeksnotities'. Beide vormen van intern onderzoek zijn uitgevoerd door medewerkers van het WODC. De betreffende onderzoeken beslaan een breed terrein en hebben betrekking op (jeugd)criminaliteit, op de evaluatie van strafrechtelijke en bestuursrechtelijke maatregelen en op het verloop van bepaalde bedrijfsprocessen (bijvoorbeeld de zaakstromen bij rechtbanken). De betreffende onderzoeken kennen ook een zekere variatie naar de gehanteerde onderzoekspopzet. Soms is er sprake van een breedopgezette ex-post-evaluatie of van een uitgebreid literatuuronderzoek, in andere gevallen wordt bestaand statistisch materiaal gebruikt of wordt de uitvoeringspraktijk (interne beleidsregels) van een rechtbank, politiekorps of bestuursorgaan in kaart gebracht. Het interne onderzoek is vaak van langere duur (meer dan een jaar). Om aan de hand van een beperkte steekproef van drie rapporten toch een goed zicht te krijgen op de kwaliteit van de onderzoeksproductie van het WODC zijn drie, wat meer omvangrijke, onderzoeken voor nadere beoordeling geselecteerd. Al deze drie onderzoeken zijn gepubliceerd in de reeks 'Onderzoek en Beleid'. Een en ander betekent dat de commissie zich geen oordeel heeft gevormd over de publicaties ten aanzien van de minder omvangrijke onderzoeken die verschijnen in de reeks 'Onderzoeksnotities'.

Wat het *externe onderzoek* betreft dat in opdracht van het WODC door derden wordt verricht, werden er in het verslagjaar 35 studies afgerond. Deze rapporten worden onder verantwoordelijkheid van het betreffende onderzoeksinstituut (de opdrachtnemer) gepubliceerd. Ook hier treffen wij een aanzienlijke variëteit aan soorten onderzoek. Naast

evaluatie-onderzoeken van bestaand beleid en diverse literatuurstudies trof de commissie enquête-onderzoek, veldstudies, scenariostudies ten behoeve van de mogelijkheden voor toekomstig beleid en diverse probleeminventarisaties aan. Ook het uitbestede onderzoek beweegt zich over het brede terrein van het recht. Over het algemeen gaat het bij het externe onderzoek om kortlopende contracten (onderzoeksduur kleiner dan een jaar).

De zeven geselecteerde onderzoeken zijn beoordeeld op wetenschappelijke kwaliteit en bruikbaarheid voor het beleid. Het aspect van de *wetenschappelijke kwaliteit* is zowel gezien vanuit een methodologisch gezichtspunt ('zit het onderzoek goed in elkaar?') als vanuit een meer inhoudelijk gezichtspunt ('wordt er een goed gebruik gemaakt van de inmiddels in het betreffende vakgebied verworven en geaccepteerde inzichten?'). Ten aanzien van de *bruikbaarheid voor het beleid* is zowel gelet op de aansluiting tussen de *beleidsvraag* ten departemente en de probleemstelling van de onderzoeker als op de inhoud van de eindconclusies (eventueel ook de acceptatie door het veld van de *aanbevelingen*).

De werkwijze van de commissie is geweest dat de zeven rapporten eerst afzonderlijk werden beoordeeld door twee commissieleden. Na de beoordeling van de afzonderlijke rapporten zijn er enkele algemene conclusies en aandachtspunten geformuleerd. Bij het vormen van een algemeen oordeel is vooral gebruik gemaakt van een globale vergelijking van het interne onderzoek met het externe onderzoek.

Het voorliggende rapport is in beginsel openbaar. Uiteraard heeft het voorliggende rapport beperkingen. De belangrijkste beperking is dat slechts zeven onderzoeken zijn beoordeeld op een totale jaarlijkse onderzoeksoutput van 50 rapporten. Een tweede beperking is dat van het interne onderzoek vooral het meer omvangrijke deel is bekeken, terwijl van het externe onderzoek naast twee omvangrijke (en officieel gepubliceerde) onderzoeken, ook twee wat kleinere (niet extern gepubliceerde) onderzoeken zijn bestudeerd.

De audit

Het valt de commissie op dat *het interne, door het WODC zelf verrichte, onderzoek* gekenmerkt wordt door een streven naar degelijkheid.

- Alle drie beoordeelde onderzoeken raken de kern van het justitiële beleidsveld.
- Elk onderzoek werd begeleid door een breed samengestelde begeleidingsgroep waarin de nodige (ook externe) kwaliteit vertegenwoordigd is.
- Aan het onderzoek en de publicaties is alle zorg besteed. Het zijn goed leesbare rapporten die toegankelijk zijn voor lezers uit het juridische beleidsveld.
- De onderzoekers hebben de tijd genomen om de probleemstelling uit te werken op basis van eerder, ander onderzoek (literatuur). De onderzoeksopzet is doordacht, al is de commissie het niet altijd met de gemaakte keuzen eens.
- Over het algemeen hebben de interne WODC-onderzoeken een (aanmerkelijk) ruimer tijdsbudget gekregen dan de externe WODC-onderzoeken. Dit ruimere tijdsbudget wordt door de interne onderzoekers overigens wel eens overschreden.
- Het valt de commissie op dat de interne WODC-onderzoekers een betrekkelijk kritische houding ten toon spreiden tot het officiële beleid. Er wordt soms afstand genomen van de oorspronkelijke opdracht (probleemstelling), maar de van de oorspronkelijke opzet afwijkende beleidsaanbevelingen zijn goed onderbouwd en blijken overtuigend te werken in het beleidsveld. Deze onbevagen houding ten opzichte van het beleid verdient waardering.
- Verder valt het op dat het WODC ook de meer kwantitatieve onderzoeken niet uit de weg

gaat. De bewerking van bestaand statistisch materiaal biedt inderdaad veel mogelijkheden om beleidsvragen en onderzoeksvragen te beantwoorden. Hier ligt bij uitstek ook een taak voor de in het beleidsveld goed ingevoerde onderzoekers van het WODC. Wel is het opvallend dat hier en daar minder geavanceerde analyse-technieken worden gebruikt dan wenselijk is.

- Het verdient aanbeveling om zowel bij de opzet als bij de eerste interne presentatie van de gegevens een methodologische toets te laten uitvoeren door een statisticus. Een en ander komt overeen met kwaliteitseisen zoals deze worden gesteld in de ISO-normering van wetenschappelijk onderzoek.

Ten aanzien van *het externe, door het WODC uitbestede onderzoek* moet worden opgemerkt dat slechts de helft hiervan als een extern rapport is gepubliceerd, terwijl de andere helft meer valt in de rubriek interne onderzoeksnotities (grijze publicaties). Aan deze interne onderzoeksnotities wordt geen ruime bekendheid gegeven. Gezien de inhoud (en/of de kwaliteit) van de rapporten die de commissie onder ogen heeft gekregen valt dit ook te verdedigen.

- Over het algemeen blijven de uitbestede onderzoeken dicht bij de oorspronkelijke probleemstelling. De houding van de externe onderzoekers ten opzichte van de beleidsvraag is volgend en niet kritisch.

- De gekozen onderzoeksvormen zijn degelijk en adequaat, maar weinig vernieuwend. Het gaat vaak om standaard-onderzoek.

- Wel moet hieraan worden toegevoegd dat het externe onderzoek moet plaatsvinden onder een (soms zeer) grote tijdsdruk. Dit laatste komt de kwaliteit van de onderzoeksresultaten waarschijnlijk niet ten goede.

- Opvallend zijn de gebreken die de externe rapporten vertonen in de behandeling van de algemene literatuur, ander onderzoek en de bespreking van de beleidscontext. Het verdient aanbeveling om de externe onderzoekers iets meer tijd te geven (om een goed rapport te schrijven) en er iets sterker van te doordringen dat ze deels schrijven voor een publiek van geïnteresseerde juristen en beleidsmedewerkers in het beleidsveld.

- Opvallend is dat sommige externe onderzoeken op de rand van (of misschien wel buiten) het justitiële beleidsveld in enge zin liggen en wat dat betreft ook onder de verantwoordelijkheid van andere departementen hadden kunnen worden uitbesteed (en begeleid).

Groningen, Utrecht, Nijmegen, Omaha
28-8-99

EEN EXTERNE BEOORDELING VAN IN 1998 GEPUBLICEERD WODC-ONDERZOEK: HET MANAGEMENTRAPPORT

Michiel Herweijer, Peter van der Heijden, Jan Leijten, Ineke Haen.

I Inleiding

Om de Adviesraad van het WODC in staat te stellen zich een oordeel te vormen over de kwaliteit van het onderzoek dat onder de verantwoordelijkheid van het WODC naar buiten wordt gebracht, is door de directeur van het WODC, prof.dr.H.G. van de Bunt, op 12 maart 1999 een onafhankelijke visitatiecommissie ingesteld. Het was de bedoeling dat deze externe commissie haar oordeel over het WODC-onderzoek en de eventuele suggesties tot verbetering voor 1 september 1999 zou voorleggen aan de Adviesraad van het WODC.

De commissie van externe beoordelaars werd gevormd door vier leden: prof.dr.I. Haen (hoogleraar criminologie, University of Nebraska), prof.dr.P.G.M. van der Heijden, hoogleraar statistiek ten behoeve van de sociale wetenschappen, Universiteit van Utrecht), prof.dr.M. Herweijer (hoogleraar bestuurskunde, Rijksuniversiteit Groningen) en prof.dr.J. Leijten (emeritus hoogleraar civiel procesrecht Nijmegen en Advocaat-Generaal bij de Hoge Raad). De commissie is in de periode mei-augustus 1999 drie keer bijeen geweest.

Om tot een kwaliteitsoordeel te komen heeft de commissie een kleine greep gedaan uit het totale aanbod van het in 1998 afgeronde WODC-onderzoek. Bij de selectie van nader te beoordelen onderzoeken is een onderscheid gemaakt tussen 'interne onderzoeken' en 'externe onderzoeken'. Ten aanzien van deze selectie heeft prof.dr.H.G. van de Bunt geadviseerd.

Uit het Jaarverslag 1998 van het WODC kwam naar voren dat er in dit verslagjaar 15 *interne onderzoeken* werden afgerond. Acht daarvan (de meer omvangrijke onderzoeken) werden gepubliceerd in de reeks 'Onderzoek en Beleid'. Zeven daarvan (de minder omvangrijke) werden gepubliceerd in de reeks 'Onderzoeksnotities'. Beide vormen van intern onderzoek zijn uitgevoerd door medewerkers van het WODC. De betreffende onderzoeken beslaan een breed terrein en hebben betrekking op (jeugd)criminaliteit, op de evaluatie van strafrechtelijke en bestuursrechtelijke maatregelen en op het verloop van bepaalde bedrijfsprocessen (bijvoorbeeld de zaakstromen bij rechtbanken). De betreffende onderzoeken kennen ook een zekere variatie naar gehanteerde onderzoeksopzet. Soms is er sprake van een breedopgezette ex-post-evaluatie of van een uitgebreid literatuuronderzoek, in andere gevallen wordt bestaand statistisch materiaal gebruikt of wordt de uitvoeringspraktijk (interne beleidsregels) van een rechtbank, politiekorps of bestuursorgaan in kaart gebracht. Het interne onderzoek is vaak van langere duur (meer dan een jaar).

Om aan de hand van een beperkte steekproef van drie rapporten toch een goed zicht te krijgen op de kwaliteit van de onderzoeksproductie van het WODC zijn drie, wat meer omvangrijke, onderzoeken voor nadere beoordeling geselecteerd. Al deze drie onderzoeken zijn gepubliceerd in de reeks 'Onderzoek en Beleid'. Een en ander betekent dat de commissie zich geen oordeel heeft kunnen vormen over de publicaties die verschijnen in de reeks 'Onderzoeksnotities'.

De volgende rapporten werden beoordeeld: J.M. Nelen en V. Sabee, 'Het vermogen te ontnemen: een evaluatie van de ontnemingswetgeving'. Het gaat hier om uitgebreide ex-post-evaluatie (zie de bijlagen 1 en 2). Het tweede interne onderzoek dat werd geselecteerd is: R.J.J. Eshuis, 'Een kwestie van tijd: Een onderzoek naar de doorlooptijden in handelszaken'

(zie bijlage 3). Ook hier kan worden gesproken van een uitgebreide ex-post-evaluatie. En tenslotte: M.W. Bol, G.J. Terlouw, L.W. Brees, C. Verwers, 'Jong en gewelddadig: Ontwikkeling en achtergronden van de geweldscriminaliteit onder jeugdigen' (zie: bijlage 4). In dit geval kan gesproken worden van een combinatie van literatuurverkenning, bewerking van bestaand statistisch materiaal en interviews met sleutelinformanten.

Wat het *externe onderzoek* betreft dat in opdracht van het WODC door derden wordt verricht, werden er in het verslagjaar 35 studies afgerond. Deze rapporten worden onder verantwoordelijkheid van het betreffende onderzoeksinstituut (de opdrachtnemer) gepubliceerd. Ook hier trof de commissie een aanzienlijke variëteit aan soorten onderzoek. Naast evaluatie-onderzoeken van bestaand beleid en diverse literatuurstudies, gaat het hier om enquête-onderzoek, veldstudies, scenariostudies ten behoeve van de mogelijkheden voor toekomstig beleid en diverse probleeminventarisaties. Ook het uitbestede onderzoek beweegt zich over het brede terrein van het recht. Over het algemeen gaat het bij het externe onderzoek om kortlopende contracten (onderzoeksduur kleiner dan een jaar).

Uit deze verzameling extern onderzoek werden vier onderzoeken ter beoordeling geselecteerd. Twee kleinere onderzoeken en twee meer omvangrijke onderzoeken. De twee kleinere externe onderzoeken zijn: J.H. Kerstholt en A.G. van Amelsvoort, 'Een vergelijking van confrontatiemethoden: Oslo, video en foto's' (zie de bijlagen 5 en 6) en Oranjewoud, 'Vergunningverlening bij bedrijfsvestiging' (zie bijlage 7). In het eerste geval gaat het om een laboratoriumexperiment waarin een aantal methoden van identificatie worden vergeleken. In het tweede geval gaat om een probleem-inventariserend gevalsonderzoek.

De tweede grotere, externe onderzoeken hebben beide betrekking op de evaluatie van de Politiewet 1993. In de eerste plaats werd het ITS-rapport L. Gunther Moor, L. Bakker en I. Brummelkamp, 'Evaluatie Politiewet 1993: Een breedte-onderzoek' bestudeerd (zie: bijlage 8). Het betreft hier een primair beschrijvende ex-post-evaluatie die is uitgevoerd op basis van bestaande documenten en telefonische ondervraging van ruim honderd sleutelinformanten. In de tweede plaats ging het om: U. Rosenthal, G.J.N Bruinsma, e.a., 'Evaluatie Politiewet: Een diepte-onderzoek' (zie: bijlage 9). Dit onderzoek werd verricht door het 'universitair consortium politie-onderzoek'. Het gaat hier om een ex-post-evaluatie op basis van een zestal cases.

De zeven geselecteerde onderzoeken zijn beoordeeld op wetenschappelijke kwaliteit en bruikbaarheid voor het beleid. Het aspect van de *wetenschappelijke kwaliteit* is zowel gezien vanuit een methodologisch gezichtspunt ('zit het onderzoek goed in elkaar?') als vanuit een meer inhoudelijk gezichtspunt ('wordt er een goed gebruik gemaakt van de inmiddels in het betreffende vakgebied verworven en geaccepteerde inzichten?'). Ten aanzien van de *bruikbaarheid voor het beleid* is zowel gelet op de aansluiting tussen de *beleidsvraag* ten departemente en de probleemstelling van de onderzoeker als op de inhoud van de eindconclusies (eventueel ook de acceptatie van de *aanbevelingen*).

In het onderstaande worden de zeven rapporten eerst afzonderlijk beoordeeld. Meer gedetailleerde beoordelingen zijn opgenomen in de negen bijlagen. Van twee rapporten zijn er in de bijlagen twee beoordelingen opgenomen. Na de beoordeling van de afzonderlijke rapporten volgen er enkele algemene conclusies. Bij het vormen van dit algemene oordeel is vooral gebruik gemaakt van een globale vergelijking van het interne onderzoek met het externe onderzoek.

Het voorliggende rapport heeft een *vertrouwelijk* karakter. Het rapport wordt alleen ter beschikking gesteld aan het management van het WODC en de Adviesraad. Wel kunnen de WODC-medewerkers die het aangaat kennis nemen van dat deel van het voorliggende rapport

(en de betreffende bijlage) dat betrekking heeft op het door hen uitgevoerde dan wel begeleidde onderzoek.

Uiteraard heeft het voorliggende rapport beperkingen. De belangrijkste beperking is dat slechts zeven onderzoeken zijn beoordeeld op een totale jaarlijkse onderzoeksoutput van 50 rapporten. Een tweede beperking is dat van het interne onderzoek vooral het meer omvangrijke deel is bekeken, terwijl van het externe onderzoek zowel twee wat grotere (officieel gepubliceerde) als twee wat kleinere (niet extern gepubliceerde) rapporten zijn bestudeerd.

II Beoordeling van de interne onderzoeken

'HET VERMOGEN TE ONTNEMEN' (zie de bijlagen 1 en 2)

- het gaat hier om een goed leesbaar rapport, met een heldere probleemstelling en een duidelijke uitwerking van de vraagstelling;
- de nonrespons is een probleem dat meer aandacht had mogen krijgen;
- over de lopende zaken die nog niet zijn afgerond, wordt geen nadere informatie verstrekt; dat laatste zou het beeld meer compleet gemaakt hebben;
- de kwantitatieve analyse is in beginsel adequaat (hoewel de lopende zaken er nog buiten zijn gebleven);
- de kwalitatieve analyse is adequaat, en geeft een verrijking van het inzicht;
- de conclusies zijn onderbouwd;
- in de begeleidingscommissie ontbreekt een methodoloog op het gebied van evaluatie-onderzoek;
- aan het onderzoek is op een adequate wijze bekendheid gegeven; de aanbevelingen zijn beleidsrelevant en mogen rekenen op acceptatie in het veld.

'EEN KWESTIE VAN TIJD' (zie bijlage 3)

- in deze beleidsevaluatie is terecht gebruik gemaakt van 'rolgegevens' waardoor een vergelijkbare voor- en nameting gerealiseerd kon worden;
- de steekproef van 17.000 zaken is te omvangrijk geweest. Ook met een kleinere steekproef waren significante uitspraken mogelijk geweest. Hiertoe had vooraf een zogeheten 'poweranalyse' uitgevoerd kunnen worden;
- tussen de periode van de voormeting en de periode van de nameting zit slechts een korte periode. Onvoldoende maken de onderzoekers duidelijk in hoeverre de rechtbanken er in zijn geslaagd de experiment-variabele ('versneld regiem') ook daadwerkelijk in de praktijk te brengen zoals dat volgens de regels van de kunst zou moeten;
- bij de analyse van verschillen wordt voornamelijk gebruik gemaakt van gemiddelden. De doorlooptijd is echter een scheef verdeelde variabele. Aan statistische significantie wordt weinig tot geen aandacht besteed. Bovendien zijn niet van alle zaken de doorlooptijden voltooid. Een survival-analyse had hier behulpzaam kunnen zijn. De gehanteerde analysetechnieken zijn onvoldoende geavanceerd;
- de oorspronkelijke effectiviteitsvraag wordt in dit onderzoek niet van een gekwantificeerd antwoord voorzien. De geconstateerde verschillen in de doorlooptijd worden niet nader geanalyseerd met behulp van regressietechnieken teneinde de samenhang met achterliggende of doorkruisende rechtbankkenmerken (zoals werkvoorraad, werkaanbod) vast te stellen en de netto-bijdrage van de experimentele maatregel vast te stellen;

- in het rapport worden de beleidsdiscussies (voors en tegens van versneld regiem), de gekozen onderzoeksopzet, de werkwijze, de zwakheden van de verzamelde gegevens helder en controleerbaar uiteengezet;
- in de aanbevelingen treden de onderzoekers buiten de oorspronkelijke probleemstelling. Onduidelijk is overigens of deze beleidsaanbevelingen (bijvoorbeeld over de (de)centrale 'planning en control' van rechtbanken) voor rekening komen van de onderzoekers, dan wel afkomstig zijn van de externe begeleidingscommissie;
- aan het onderzoek is terecht en op correcte wijze de nodige bekendheid gegeven. Het onderzoek heeft veel gegevens opgeleverd (ook buiten de oorspronkelijke vraagstelling om). Het rapport is goed leesbaar en er wordt een goed gebruik gemaakt van de beschikbare literatuur.

'JONG EN GEWELDDADIG' (zie bijlage 4)

- het cijfermatige deelonderzoek is waardevol, controleerbaar en informatief;
- de kanttekeningen (nuanceringen) bij het cijfermatige onderzoek zijn niet op alle punten verhelderend;
- de literatuurstudie heeft geen duidelijke vraagstelling en legt zich onnodige beperkingen op;
- onduidelijk is hoe de literatuurselectie werd uitgevoerd; er worden ook bronnen gebruikt die niet zijn gebaseerd op oorspronkelijke data;
- de literatuurstudie is meer het verslag van een zoektocht dan een synthese van inmiddels bereikte en geaccepteerde inzichten;
- de verrijking van het onderzoek met meningen van daders en oordelen van deskundigen maakt een weinig overtuigende indruk; er is sprake van een journalistieke benadering; voor de handliggende bezwaren ten aanzien van de representativiteit van het opgetekende materiaal worden niet weggenomen;
- de conclusies zijn niet onderbouwd;
- het was beter geweest wanneer het rapport zich had beperkt tot alleen het eerste cijfermatige deel;
- het onderzoek heeft vertraging opgelopen.

III Beoordeling van de externe onderzoeken

'EEN VERGELIJKING VAN CONFRONTATIEMETHODEN' (zie de bijlagen 5 en 6)

- terecht is in dit geval gekozen voor het klassieke experiment;
- de aansluiting met eerder experimenteel onderzoek naar identificatiemethoden is in het rapport onvoldoende gedocumenteerd;
- de theoretische toelichting op het experimentele onderzoek is bepaald summier;
- de conclusies zijn onderbouwd;
- het onderzoek voldoet aan de binnen het vakgebied geldende (hoge) eisen;
- de selectie van de proefpersonen wordt summier verantwoord; enkele vragen blijven open staan;
- de beleidsrelevantie van het onderzoek is zeker aanwezig;
- de resultaten van het onderzoek lenen zich minder goed voor een algemene (vrije) verspreiding onder bijvoorbeeld advocaten in strafzaken.

'VERGUNNINGVERLENING BIJ BEDRIJFSVESTIGING' (zie bijlage 7)

- het onderzoek heeft het WODC weinig gekost en het werd snel geleverd;
- de leesbaarheid van het rapport is gering;
- de verzamelde gegevens zijn niet representatief voor het bedrijfsleven als geheel;
- er vindt geen enkele verwijzing plaats naar ander recent onderzoek naar hetzelfde verschijnsel, in het rapport wordt geen literatuur vermeld;
- onderzoekers hebben onvoldoende kennis van het geldend recht en zijn in hun onjuiste beeldvorming terzake onvoldoende door de begeleidingscommissie gecorrigeerd;
- de probleemstelling van het onderzoek is eenzijdig en dat geldt ook voor de geraadpleegde respondenten;
- de verzamelde gegevens worden wel 'gemiddeld', maar niet nader geanalyseerd op onderlinge verbanden; op basis van deze beschrijvende-beoordelende informatie kan de stuurgroep geen pregnante beleidsaanbevelingen ontwikkelen;
- waarschijnlijk heeft het rapport geen externe bekendheid gekregen en is het alleen bedoeld geweest voor intern gebruik in de ambtelijke MDW-werkgroep.

'EVALUATIE POLITIEWET: BREEDTEONDERZOEK' (zie bijlage 8)

- dit onderzoek is onder betrekkelijk grote tijdsdruk uitgevoerd; het heeft een sterk beschrijvend en inventariserend karakter; het voldoet niet aan de stringente eisen van een effectiviteitsonderzoek;
- het onderzoek geeft op basis van documenten een adequate momentopname van de wijze waarop regionale politiekorpsen hun interne organisatie hebben ingericht;
- het onderzoek geeft op basis van telefonische interviews een representatieve indruk van de klachten die er onder de politiebesteders blijken te bestaan over het functioneren van de politiekorpsen; de verschillen tussen de korpsen en de daar heersende stijl van leiding geven komen goed naar voren;
- de wijze waarop het onderzoek is uitgevoerd wordt duidelijk uit de doeken gedaan;
- het onderzoek graaft niet erg diep; een expliciete vergelijking met de situatie voor de herziene Politiewet ontbreekt; een toerekening van geconstateerde veranderingen in het politieoptreden aan de herziening van de Politiewet vindt niet plaats; een verklaring voor gevonden verschillen tussen de politieregio's ontbreekt;
- de verslaglegging is uitstekend; het is een leesbaar rapport, waaraan terecht publiciteit is gegeven;
- de beleidsrelevantie is niet onmiddellijk duidelijk; de onderzoekers houden zich op de vlakte; er zijn geen beleidsaanbevelingen.

'EVALUATIE POLITIEWET: DIEPTEONDERZOEK' (zie bijlage 9)

- dit onderzoek is nogal prijzig wanneer men bedenkt dat slechts de organisatie van zes politieregio's in kaart is gebracht;
- het betreffende onderzoek is tijdig afgerond, hoewel er slechts enkele maanden beschikbaar waren om het onderzoek uit te voeren;
- hoewel er wordt gesproken van een kwalitatief onderzoek worden er weinig echte (historische) 'processen' (op een levendige en gedetailleerde wijze) in kaart gebracht; ten opzichte van het breedte-onderzoek heeft het diepte-onderzoek niet veel meer diepgang;
- hoewel de onderzoekers van het diepte-onderzoek normatieve oordelen (confrontatie van wet en werkelijkheid) niet uit de weg gaan, voldoet het onderzoek niet aan de stringente eisen die aan een effectiviteitsstudie gesteld mogen worden. Veranderingen in de politiekorpsen na invoering van de Politiewet worden niet toegeschreven aan rivaliserende factoren dan wel de

wetswijziging. De netto-bijdrage van de Politiewet wordt niet expliciet geconfronteerd met de wensen van de wetgever;

- het onderzoek leidt wel tot heldere uitspraken met een evaluatief karakter, maar een systematische vergelijking van de verschillen en overeenkomsten tussen de zes (vanwege hun variëteit geselecteerde) cases ontbreekt;
- het rapport is helder weergegeven en terecht is er veel aandacht besteed aan de verspreiding van de bevindingen.

'DE ORGANISATIE VAN DE BEIDE POLITIE-ONDERZOEKEN'

- achteraf kan de vraag worden gesteld of een opsplitsing van het beschikbare budget voor het kwalitatieve diepte-onderzoek in kleinere moten niet meer voordelen zou hebben geboden; door het budget van het kwalitatieve onderzoek als één geheel uit te besteden, moest een vrij zware (EU-)aanbestedingsprocedure worden gevolgd; dit leidde tot vertraging; het resultaat was hetgeen men waarschijnlijk wilde vermijden, namelijk dat drie verschillende instituten afzonderlijk casestudies hebben verricht. De drie verschillende onderzoeksgroepen die aan het kwalitatieve diepteonderzoek hebben gewerkt, hebben onder grote tijdsdruk gestaan.
- de vraag kan worden opgeworpen waarom niet eerder is begonnen met de voorbereiding van het evaluatie-onderzoek. In 1995 was al bekend dat er een evaluatie zou komen. Waarom kon de externe evaluatiecommissie (de cie-Ringeling) niet op een eerder moment worden ingesteld om alvast te beginnen met de voorbereiding (bijvoorbeeld een nulmeting)?
- de tijdsdruk en de ingewikkelde procedure hebben er wellicht aan bijgedragen dat de combinatie van het breedte- met het diepte-onderzoek een geringere toegevoegde waarde heeft gehad dan wellicht werd nagestreefd.

IV ALGEMENE OBSERVATIES: DE AUDIT

Het valt de commissie op dat *het interne, door het WODC zelf verrichte, onderzoek* gekenmerkt wordt door een streven naar degelijkheid.

Alle drie onderzoeken raken de kern van het justitiële beleidsveld.

Elk onderzoek wordt begeleid door een breed samengestelde begeleidingsgroep waarin de nodige kwaliteit vertegenwoordigd is.

Aan het onderzoek en de publicaties is alle zorg besteed. Het zijn goed leesbare rapporten die toegankelijk zijn voor lezers uit het juridische beleidsveld.

Men heeft de tijd genomen om de probleemstelling uit te werken op basis van eerder, ander onderzoek (literatuur).

De onderzoeksopzet is doordacht, hoewel de commissie het niet altijd eens is met de gemaakte keuzen.

Over het algemeen hebben de interne WODC-onderzoeken een ruimer tijdsbudget dan de externe WODC-onderzoeken. Dit ruimere tijdsbudget wordt door de interne onderzoekers overigens ook nog wel eens overschreden.

Het valt de commissie op dat de interne WODC-onderzoekers een betrekkelijk kritische houding ten toon spreiden tot het officiële beleid. Er wordt soms afstand genomen van de oorspronkelijke opdracht (probleemstelling), maar de van de oorspronkelijke opzet afwijkende beleidsaanbevelingen zijn goed onderbouwd en blijken overtuigend te werken in het beleidsveld. Deze onbevangen houding ten opzichte van het beleid verdient waardering.

Verder valt het op dat het WODC de meer kwantitatieve onderzoeken niet uit de weg gaat. De bewerking van bestaand statistisch materiaal biedt inderdaad veel mogelijkheden om beleidsvragen te beantwoorden. Hier ligt bij uitstek ook een taak voor de in het beleidsveld goed ingevoerde onderzoekers van het WODC. Wel is het opvallend dat hier en daar minder geavanceerde analyse-technieken worden gebruikt.

Het verdient aanbeveling om zowel bij de opzet van het onderzoek als bij de eerste interne presentatie van de gegevens een methodologische toets te laten uitvoeren door een statisticus. Een en ander komt overeen met kwaliteitseisen zoals deze worden gesteld in de ISO-normering van wetenschappelijk onderzoek.

Ten aanzien van *het externe, door het WODC uitbestede onderzoek* moet worden opgemerkt dat slechts de helft hiervan echt als een extern rapport is gepubliceerd, terwijl de andere helft meer valt in de rubriek interne onderzoeksnotities (grijze publicaties). Aan deze interne onderzoeksnotities wordt geen ruimere bekendheid gegeven. Gezien de inhoud van de rapporten die de commissie onder ogen heeft gekregen is dit ook terecht.

Over het algemeen blijven de uitbestede onderzoeken dicht bij de oorspronkelijke probleemstelling. De houding van de externe onderzoekers ten opzichte van de beleidsvraag is bepaald volgend en zeker niet kritisch.

De gekozen onderzoeksvormen zijn degelijk en adequaat, maar weinig vernieuwend. Het gaat vaak om binnen de betreffende discipline volkomen geaccepteerd standaard-onderzoek. Wel moet hieraan worden toegevoegd dat het externe onderzoek moet plaatsvinden onder een (soms zeer) grote tijdsdruk. Dit laatste komt de kwaliteit van de onderzoeksresultaten zeker niet ten goede.

Opvallend zijn de gebreken die de externe rapporten vertonen in de behandeling van algemene literatuur, ander onderzoek en de bespreking van de beleidscontext.

Het verdient aanbeveling om de externe onderzoekers iets meer tijd te geven (om een goed

rapport te schrijven) en iets sterker er van te doordringen dat ze deels schrijven voor een publiek van geïnteresseerde juristen in het beleidsveld.

Opvallend is dat sommige externe onderzoeken op de rand van (of misschien wel buiten) het justitiële beleidsveld in enge zin liggen. De vraag kan worden gesteld of bepaalde onderzoeken niet beter onder het beheer van een ander departement hadden kunnen worden uitgevoerd.

Groningen, Nijmegen, Utrecht, Omaha
28-8-99

Bijlage 1:

Oordeel ten aanzien van WODC 170:

HET VERMOGEN TE ONTNEMEN.

1. De ontneming van het wederrechtelijk genoten voordeel is in het Nederlandse strafproces een geheel nieuwe figuur, in zekere zin een Fremdkörper afkomstig uit het civiele recht. De strafrechtelijk veroordeelde krijgt straf en moet, dat is al oud, de schade van het slachtoffer vergoeden. Maar vaak is er geen slachtoffer of is het slachtoffer moeilijk aan te wijzen of te identificeren, terwijl de veroordeelde zich toch wederrechtelijk heeft verrijkt. Die verrijking wordt ongedaan gemaakt door hem zijn onrechtmatige winsten af te nemen ten behoeve van de Staat. Deze is niet, althans hoeft niet te zijn, degene die de schade heeft geleden en in zoverre wijkt de ontnemingsfiguur van het civiele recht af, dat niet alleen onrechtmatige verrijking bij de een vergt maar ook de daarmee corresponderende schade bij de ander. Het is duidelijk dat de figuur van de ontneming pas zin kreeg, en dus pas ingevoerd werd, toen criminaliteit een zaak van grote winsten ging betreffen, anders gezegd, toen handel in soft- en hard drugs in hoge mate winstgevend werd. Bij diefstal en inbraak is er normaliter ook onrechtmatige verrijking. Maar daarvoor had de wetgever dit instituut niet nodig. Gelet op wat de heler voor gestolen goed over heeft, zal het overgrote merendeel van dieven en inbrekers niet rijk worden. Bovendien is bij deze gevallen meestal een benadeelde aan te wijzen, wiens aanspraken voor die van de Staat gaan.

Het rapport maakt zonneklaar - en dat vind ik zeer belangrijk - dat deze nieuwe rechtsfiguur, mede door de zgn. tien-zaken norm, inhoudende dat iedere officier van justitie per jaar tien ontnemingszaken moet aanbrengen, aan zijn eigenlijke doel voorbij dreigt te gaan: rijke criminelen in hun beurs aan te tasten, doordat kleine ontnemingszaken (buit van diefstal enz.) een steeds groter gedeelte van het geheel aan ontnemingszaken gaan vormen. Voor dat soort zaken hebben we het altijd zonder de figuur van de ontneming kunnen stellen.

2. Voorts vind ik helder uiteengezet dat en vooral waarom het instituut in de onderzoeksperiode niet goed heeft gefunctioneerd. Het rapport legt er terecht de nadruk op dat het van het OM afhangt of er zaken aanhangig worden gemaakt en of dat de zaken zijn waarop deze nieuwe rechtsfiguur doelt: kennis ervan druppelt in vele terzake doende kringen maar langzaam door; het kunnen werken met de figuur wordt onder meer door geringe bereidheid tot het instellen van aparte ontnemingskamers bemoeilijkt, alsmede doordat ook bij het OM maar weinig werkkraft daarvoor overblijft en die twee samen bewerken weer dat het willen optreden als eiser in een ontnemingszaak niet altijd al te groot is. Ook is er, zoals het rapport, zij het vooral indirect aangeeft, te weinig rechtspraak in hoogste instantie, welke rechtspraak dan ook dikwijls nog ontmoedigend werkt, om met een gerust hart aan grote ontnemingsprocessen te beginnen. De drie hiervoor onderstreepte criteria, beslissend voor het wetslagen van de rechtsfiguur, zijn helder afzonderlijk uitgewerkt voor alle groepen die ermee te maken krijgen (politie, o.m. rechters en advocaten).

3. Het rapport is lang en de huidige consumenten van rapporten - en dat zijn toch in hoofdzaak juristen en de politie - hebben bijna altijd haast. Toch vind ik het rapport niet te lang. Het is goed geschreven en wie het doorneemt krijgt een goed totaalbeeld van deze moeilijke materie waarmee in de toekomst, naar te verwachten valt, in toenemende mate gewerkt zal moeten worden.

Bijlage 2:

Oordeel ten aanzien van WODC 170:

HET VERMOGEN TE ONTNEMEN

In het algemeen is het een goed rapport. Geschreven in helder Nederlands, goed leesbaar. Goede layout, en dat geldt ook voor de tabellen.

De aanwezigheid van een brede begeleidingscommissie is van belang, en het is zorgvuldig dat hierin is voorzien. Hierin ontbreekt overigens wel een methodoloog op het gebied van evaluatieonderzoek. Onduidelijk of die het onderzoek had kunnen verbeteren.

Er is sprake van een heldere probleemstelling en vraagstelling. De keuze van het design is redelijk. De voor de interviews gekozen groep is goed, maar de respons ligt bij chefs van basiseenheden ver beneden 100%. Dit is toch wel een probleem. De onderzoekers zullen dit ongetwijfeld met volledige inzet doen, maar hebben zij experts geraadpleegd die het non-respons probleem zo klein mogelijk kunnen maken?

De keuze voor kwalitatieve ondervraging is goed verdedigbaar omdat je nog niet goed weet wat eruit moet komen, en een en ander ook leidt tot vergroting van de leesbaarheid van het rapport.

Er zijn wel problemen bij de bestudering van de strafzaken:

- veel zaken lopen nog, en dit kunnen zaken met specifieke kenmerken zijn. De auteurs onderkennen dit, maar verhelpen dit probleem vervolgens niet. Maar had het niet de voorkeur verdiend ook een aantal van dergelijke lopende zaken nader te onderzoeken?
- 12 cases (in zekere zin uitgebreid tot 35) is erg laag. Is dit voldoende om een goed beeld te verkrijgen? Men stelt dat de cases illustratief zijn, maar beantwoordt een illustratie voldoende aan de doelstellingen van het onderzoek?

De analyse op zich is adequaat, maar resultaten kunnen vertekend zijn omdat nog lopende zaken niet zijn meegenomen. De onderzoekers gebruiken het programma KWALITAN, en dat is de standaard. Wij kunnen niet beoordelen of de onderzoekers de kwalitatieve analyses goed hebben verricht, maar dit is eigen aan kwalitatief onderzoek en geldt tot op zekere hoogte ook voor kwantitatief onderzoek (al heeft men bij kwalitatief onderzoek meer vrijheid dan bij kwantitatief onderzoek).

Het rapport is zorgvuldig en beantwoordt aan het doel. De non-respons is een probleem. Heeft men voldoende know-how op dit gebied? Het rapport is erg lang. Is dit vervelend voor doelgroep, of wordt die ook op een andere wijze bereikt? Hebben onderzoekers voldoende know-how op het gebied van kwalitatieve analyse?

Bijlage 3

Oordeel ten aanzien van WODC 171:

EEN KWESTIE VAN TIJD

De aanleiding tot het evaluatie-onderzoek is de invoering van een beleidsexperiment met het versneld regiem in acht van de negentien arrondissementsrechtbanken in 1996. Bij het voorbereiden van een uitspraak over de effectiviteit van deze beleidsmaatregel ter bereiking van een versnelling in het rechterlijk vonnis in handelszaken blijkt dat er ook geen goed zicht is op de doorlooptijden voorafgaand aan het experiment (geen nulmeting), in arrondissementen die niet

aan het experiment deelnemen (geen controlemeting) en in de mate waarin er in die verschillende rechtbanken ook (reeds) andere maatregelen (door de rechter in de vorm van 'comparitie na antwoord' en door de eiser in de vorm van 'verkorte termijn') worden genomen ter versnelling van de doorlooptijden (dus ongelijksoortigheid van de experiment- en controlerechtbanken).

Blijkbaar leeft al in een vroeg stadium de veronderstelling bij de onderzoekers dat de praktijk in de negentien rechtbanken nogal verschilt (mede door een verschillende omvang van het werkaanbod en door een verschillend lokaal beleid) zodat deze versturende factoren in het onderzoek betrokken moeten worden. Om de gestelde effectiviteitsvraag te beantwoorden moeten de onderzoekers de nulmeting achteraf reconstrueren en moeten ze vaststellen wat de overige relevante verschillen zijn tussen de zaken die in de acht arrondissementen onder het versneld regiem vallen en de zaken die daar niet onder vallen.

Gebruikmakend van de beschikbare mogelijkheden (automatisering deugt niet, dossieronderzoek is te omvangrijk, rolonderzoek is haalbaar, ambtelijke bronnen hebben voorkeur boven indrukken van sleutelpersonen) wordt de onderzoeksopzet adequaat aangepast aan de gegeven beleidscontext. De probleemstelling ('maakt VR verschil?') wordt benaderd door de gemiddelde doorlooptijden per rechtbank voor relevante contrastgroepen te berekenen en te vergelijken.

Waarom zijn er voor ongeveer 17.000 zaken gegevens verzameld en had dit aantal niet veel kleiner gekund? De indruk bestaat dat men met veel minder zaken had toegekund, bijvoorbeeld slechts met 2.000. Dit had men voor de aanvang van het onderzoek kunnen uitrekenen (door middel van zogeheten poweranalyse).

Het is opvallend dat de omvang van de waargenomen verschillen (die aanzienlijk blijken te zijn) niet nader wordt beoordeeld op statistische significantie en relatieve sterkte. Het blijft bij het vergelijken van de gemiddelde (mediane, percentiele) verschillen in doorlooptijd zonder de interactie met de achterliggende groepskenmerken (rechtbank, gebruik andere versnellingsmaatregelen, werkvoorraad) te kwantificeren. Een regressie-analyse had hier tot de mogelijkheden behoord. Een mogelijk gevolg is dat 'harde' (expliciete) antwoorden op de onderliggende effectiviteitsvragen ontbreken.

De analyse is louter beschrijvend, en van eenvoudig niveau (bijvoorbeeld bij gemiddelden worden de standaardfouten van gemiddelden en de standaarddeviaties al niet meer genoemd). Er wordt door de onderzoekers rekening mee gehouden dat de variabele 'doorlooptijd' scheef is verdeeld, waardoor het louter rapporteren van gemiddelden niet adequaat zou zijn. Een belangrijke vraag is hier waarom in het gehele rapport de analyse niet wat geavanceerder is aangepakt door gebruik te maken van modellen uit de survivalanalyse (nodig omdat (i) je niet gewoon de variabele tijd als afhankelijke variabele moet nemen omdat deze niet normaal verdeeld is; (ii) er sprake is van censurering in RT-zaken: bij veel zaken weet men niet wat de doorlooptijd is op het moment van dataverzameling, en survivalanalyse corrigeert hiervoor). Het gebruik van survivalanalyse zou als voordeel hebben gehad dat men had kunnen onderzoeken welke kenmerken van zaken leiden tot een langere doorlooptijd (i.e. wat veroorzaakt vertraging), waarbij gecorrigeerd is voor het feit dat verschillende kenmerken niet onafhankelijk van elkaar zijn. Op dit type gegevens zou men dan beleid kunnen voeren.

- Bijlage 4, p. 1, in tegenstelling tot wat wordt beweerd geven kleine aantallen zaken wel degelijk onbetrouwbare gemiddelden en staan niet toe te generaliseren.
- bijlage 6, p. 1, in tegenstelling tot wat wordt beweerd zijn bij een respons van 25% de resultaten beslist niet valide, omdat je er van uit kunt gaan dat de non-respons selectief is (hier worden betrouwbaarheid en validiteit door elkaar gehaald).
- er is onderbenutting van data doordat er geen regressieachtige technieken zijn uitgevoerd. -

het probleem van de non-respons bij enquêtes is niet helemaal goed gehanteerd.

De oorspronkelijke doelstelling ('geef informatie over de effecten van versneld regiem') wordt verbreed (gebeurt al in de onderzoeksopzet). In het eindrapport wordt de beleidsmakers informatie verschaft over een veel breder scala van mogelijkheden om de doorlooptijd van rechtbanken in handelszaken te vergroten. Vooral aan het eind van het rapport (wanneer het op het formuleren van beleidsaanbevelingen aankomt) nemen de onderzoekers de vrijheid om (kritische) uitspraken te doen over het onderliggende sturingsmodel ('centraal en uniform rolbeleid versus decentrale/lokale gebruikmaking van faciliterende mogelijkheden' en 'liever zelfevaluatie door rechtbanken dan afleggen van externe verantwoording aan hoger niveau'), over het aangrijpingspunt van het beleid ('prioriteit aan werkvoorraad wegwerken boven verdere diffusie van een effectief gebleken maatregel'), over alternatieve beleidsmaatregelen ('comparitie na antwoord' en 'verkorte termijn'). Deze aanbevelingen zijn terzake en verstandig, maar liggen niet in het verlengde van de oorspronkelijke probleemstelling ('is het verkorte regiem nu effectief of niet?'). Deze kritische distantie tot het 'beleid' is geen probleem, maar de oorspronkelijke effectiviteitsvraag wordt niet van een gekwantificeerd antwoord voorzien. Eigenlijk is de overblijvende beleidsconclusie: in andere gebieden dan de experimentele regio's zijn de problemen groter en daar moet het beleid zich voortaan op richten; de werkvoorraad aanpakken is urgenter dan de instroom versneld aan te pakken; en ook de twee bestaande vormen van versnelling kunnen mits enthousiast benut een versnelling opleveren (hoewel minder dan met het verkort regiem).

Gelet op de korte tijdsspanne na invoering van de experimentele maatregel had een kwalitatieve procesanalyse van de intermediaire processen ('hoe werkt het uit'; en 'gaat het nu ook anders') misschien meer voor de hand gelegen dan een summatieve productevaluatie op de einddoelen (waarvoor nu is gekozen waardoor mogelijk te prematuur wordt geconcludeerd).

De nulmeting (laatste helft 1995) en de experimentele meting (laatste helft 1996) liggen erg dicht tegen elkaar aan; er zit gemiddeld slechts een heel jaar - en wat de buitengrenzen betreft slechts een half jaar - tussen de nulmeting en nameting; vaak duurt het enige tijd voordat een formele maatregel ook echt is ingevoerd en voordat de praktijk zich daar naar heeft kunnen voegen; waarschijnlijk neemt de efficiency van versneld regiem na verloop van tijd verder toe.

De waarnemingsperiode van de nameting (zowel in de versneld regiemgroep als in het totale bestand afgehandelde zaken) is betrekkelijk kort; voor de nog niet-afgehandelde gevallen moet de doorlooptijd statistisch worden geschat; maar hoe kan dat als er slechts zo kort ervaring is met het verkorte regiem;

Het onderzoek is sterk ontworpen op het systeemniveau van de rechtbank als zodanig en minder op het niveau van de verschillende zaakstromen (processen) zelf; de intermediaire causale processen tussen de keuze voor versneld regiem en de uitspraak van de rechter (zoals bijvoorbeeld betere vordering, vaker schikking na comparitie, makkelijker kunnen overslaan van processtappen dan in reguliere gevallen) worden niet in een kwalitatieve procesanalyse betrokken; deze kritische kanttekening moet echter worden genuanceerd; wel wordt immers de comparitie van partijen in versneld regiem gevallen geobserveerd, deze waarnemingen lijken er overigens op te duiden dat de betreffende onderliggende veronderstelling in de beleidstheorie achter versneld regiem steun vindt (bij versneld regiem zijn de comparities inhoudelijker).

Er zijn overigens veel punten van grote waardering:

- de keuze van het design is slagvaardig aangepakt: er is gekozen voor retrospectief design (kijken naar alle zaken die tussen 1994-1996 zijn afgehandeld), en voor prospectief design (kijken bij RT naar alle zaken die in 1996 zijn aangemeld).

- door het gehele rapport heen wordt zorgvuldig melding gemaakt van alle mogelijke problemen in de dataverzameling, en van de interpretatieproblemen waar dit toe leidt.
- het onderzoek wordt in rapport en bijlagen goed gedocumenteerd;
- de structuur van het rapport is duidelijk;
- bij de opzet van het experimentele onderzoek is creatief gebruik gemaakt van de mogelijkheden die het onderzoeksveld biedt (boud);
- de lezer wordt gewezen op de beperkingen van de kwaliteitsoordelen van de responderende rechters en advocaten; op de beperkingen van de doelmatigheidsmeting; op de beperkingen van de beschikbare ambtelijke statistiek;
- terecht wordt er de nodige aandacht besteed aan verstorende factoren die per rechtbank en soms ook per zaak (bijvoorbeeld 'enthousaste en ervaren rechters trekken de versneld regiemzaken') het recht-toe-recht-aan gebruik van een experimentele meting belemmeren;
- er wordt in het empirische onderzoek terecht veel aandacht besteed aan mogelijke neveneffecten van de maatregel;
- het onderzoek is voorzien van een deskundige externe begeleidingscommissie;
- het onderzoek wordt op verschillende wijzen in de openbaarheid gebracht;
- er wordt in de theoretische opzet aangesloten bij relevante organisatiekundige literatuur (hoewel de betekenis van automatisering voor organisatieprocessen wordt gemist).

Bijlage 4:

Oordeel ten aanzien van WODC 174:

JONG EN GEWELDDADIG

Het gaat hier om een rapport van matige kwaliteit. Alleen het eerste deelonderzoek komt over als waardevol. Het WODC had er beter aan gedaan het antwoord op de kamervraag te beperken tot deelonderzoek 1.

Het tweede deel (het literatuuronderzoek) is onrijp en slecht uitgevoerd, de delen 3 en 4 overstijgen niet echt het niveau van journalistiek.

De begeleidingscommissie bestaat uit vijf mensen uit het veld en twee criminologen. Meer ervaringsdeskundigen dan wetenschappelijke deskundigen. Een methodoloog of statisticus ontbreekt, wat jammer is gezien het belang van het statistisch cijfermateriaal voor dit rapport.

De samenvatting weerspiegelt de inhoud van het rapport redelijk goed, hoewel de conclusies wat minder genuanceerd overkomen.

I.
De onderzoekers beperken zich tot delicten waaraan fysiek geweld te pas komt, *gericht tegen personen*. Ze introduceren echter ook het begrip '*antisociaal gedrag*' ("gedrag waarvan de pleger als oogmerk heeft er zelf 'beter' van te worden ten koste van een of meer anderen"). Waarom deze term geïntroduceerd wordt, is niet duidelijk - er is geen verwijzing naar een mogelijke (wetenschappelijk geaccepteerde) bron voor deze definitie, en bovendien lijkt deze definitie geen rechtstreeks verband te leggen met geweld (dat immers vaak expressief van karakter is, en waar overwegingen of men er zelf "beter" van wordt geen grote rol lijken te spelen).

II Eerste deelonderzoek: Cijfermateriaal (Trendstudie)

Inhoudelijk interessant. Alle soorten van geweld stijgen. Deze conclusie samen met de reeks kanttekeningen bij de statistieken had op zich al een leuk rapport opgeleverd.

Een paar opmerkingen. De figuren en tabellen zijn duidelijk. Echter, de beschrijvingen van de trends lezen moeizaam en zullen de doorsnee-lezer de nodige hoofdbreken kosten. Hoewel het trekken van conclusies omtrent trends uit statistieken altijd een moeilijke zaak is welke zich niet in een paar eenvoudige zinnen laat samenvatten, zou het rapport toch aan duidelijkheid gewonnen hebben als de onderzoekers hun bevindingen strakker en eenduidiger geformuleerd hadden. De beschrijving maakt nu een onnodig rommelige en gehaaste indruk. Een samenhangend punt is dat -teneinde te kunnen concluderen dat de stijging van jeugdgeweld disproportioneel is - men deze stijging moet vergelijken met de stijging van het geweld gepleegd door volwassenen. Dit is belangrijk als men wil zien of de toename in (jeugd)geweld te wijten is aan factoren die specifiek jeugdgericht zijn, of dat de toename in jeugdgeweld een onderdeel is van de algemene toename van geweld in de maatschappij. Dit gebeurt praktisch nergens in dit rapport.

De beschrijving van de beperkingen van de politieke statistieken en de self-reportgevens is goed. De onderzoekers verwijzen hier enkel naar Nederlandse publicaties op dit gebied; echter, ook in de buitenlandse literatuur is er veel op dit gebied gepubliceerd.

De onderzoekers vervallen in de fout dat zij bij het trekken van conclusies eerdere beperkingen van hun materiaal vergeten: in conclusie sectie 2.6 staat dat zij veronderstellen dat er wel een stijging is maar dat die minder groot is dan uit de statistieken blijkt. Zij geven aan dat zij dit niet kunnen staven. Twee opmerkingen hierover: (i) hoe waardevol zijn dergelijke indrukken? Dit is toch nauwelijks meer dan een gevoel. De onderzoekers geven geen overtuigend argument waarom de stijging juist minder groot - en niet juist groter - zou zijn dan die welke de cijfers laten zien; (ii) in slotconclusie paragraaf 6 is de kanttekening dat zij dit niet kunnen staven verdwenen. Op zo'n essentieel punt mag deze fout niet gemaakt worden.

III Tweede deelonderzoek: het literatuuronderzoek

Dit is een zwak hoofdstuk. Dit heeft een aantal redenen. In de eerste plaats is er geen duidelijke vraagstelling, er zijn geen subvraagstellingen. Zoals de onderzoekers zelf duidelijk maken, gaat het hoofdstuk nu meer over de oorzaken van criminaliteit, dan over de toename van de jeugdcriminaliteit. Hierdoor sluit het hoofdstuk niet goed aan bij de zojuist besproken trendstudie over het jeugdgeweld. Een mogelijke deelvraag zou zijn geweest: welke verklaringen worden er in de literatuur gevonden voor het *toenemen* van criminaliteit en geweld in het bijzonder. Bij deze vraag alleen zou een groot aantal artikelen geschrappt zijn, en misschien andere gevonden. De onderzoekers vermelden een paar keer dat ze veronderstellen dat bepaalde combinaties van risicofactoren zouden kunnen leiden tot verschillende *soorten* van geweld of *typen* van geweldplegers - deze subvraagstelling wordt echter in het rapport vrijwel niet expliciet besproken of onderzocht, behalve de opmerking dat de literatuur "in grote lijnen" deze vooronderstelling leek te bevestigen (p. 131).

In de tweede plaats kiezen de onderzoekers voor het *levensloopmodel* als het theoretisch kader voor hun literatuuronderzoek. Dit is geen goed idee. Het is waar dat het levensloopmodel momenteel een populaire benaderingswijze binnen de criminologie is; echter, de keuze werkt onnodig beperkend, het literatuuronderzoek slaat hierdoor belangrijke delen van de relevante sociaal-wetenschappelijke kennis op het gebied van geweld over. Zo heeft men bijvoorbeeld geen aandacht geschonken aan het gelegenheidsperspectief en de situationele aanpak (nadruk op de context) die gepresenteerd wordt in het werk van Cohen en Felson, Fagan e.a. [Wel is er een korte bespreking van wapenbezit en alcohol, maar deze factoren worden niet in een theoretische context geplaatst.] En dat terwijl hier ons inziens een van de beste verklaringen is te vinden voor een eventuele toename in criminaliteit: Men kan met de beste wil van de wereld niet zeggen dat dit literatuur onderzoek de *state of the art* van wetenschappelijke kennis op het gebied van

(jeugd)geweld vertegenwoordigt. De Amerikaanse literatuur op het gebied van geweld is uitgebreid - er wordt in dit rapport onvoldoende aandacht aan de buitenlandse bijdragen tot de kennis van jeugdgeweld besteed.

In de derde plaats is het literatuuronderzoek niet gedegen uitgevoerd. Het is niet duidelijk hoe de literatuurselectie heeft plaatsgevonden. Dit is mede het gevolg van het ontbreken van een duidelijke vraagstelling. Het is beter om je in een literatuuronderzoek te beperken tot artikelen met oorspronkelijke data. Anders loop je het gevaar verkeerd getrokken conclusies na te praten. Zo leunt het onderzoek nu erg op één lezing (!) van Hermanns.

In de vierde plaats is het hoofdstuk niet erg uitgerijpt. Sociologische verklaringen worden al te gemakkelijk als niet relevant afgedaan. De onderzoekers lijken zich niet te realiseren dat de huidige aandacht voor het individu als oorzaak van criminaliteit niet voortkomt uit het feit dat sociologie als *verklaring* heeft gefaald, maar mogelijk wel als *oplossing* (het is moeilijker er preventie op te baseren). Hier en daar komt het literatuuronderzoek ook over als een omgevallen boekenkast. Veel artikelen worden achter elkaar stuk voor stuk samengevat en niet geïntegreerd. Dit geldt vooral voor sectie 3.7. De paragraaf over persoonlijkheid (3.3.3) is niet kritisch genoeg geschreven - de theorie van Eysenck is door empirisch onderzoek herhaaldelijk onderuit gehaald - hiervan wordt geen melding gemaakt. Hetzelfde geldt voor de wijze waarop het begrip psychopathie (p.56) gebruikt wordt. Dit begrip is door vrijwel alle *mainstream* sociale wetenschappers afgedankt als een tautologisch en onbruikbaar 'conceptual wastebasket' begrip - een kritiek die in het huidige rapport helemaal niet aan de orde komt. De onderzoekers vermelden dat adoptiekinderen nogal eens met politie en justitie in aanraking komen, echter zonder vermelding van de bron. De nadruk op MAOA (staat zelfs in samenvatting!) is onevenwichtig en speculatief. De relevantie van het voorbeeld van vijf mannen voor het toegenomen geweld in Nederland ontgaat ons

In de vijfde plaats is er onvoldoende terugkoppeling naar de oorspronkelijke vraagstelling met betrekking tot de eventuele stijging van het geweld.

Onze conclusie is dat de auteurs zich vertild hebben aan de opdracht. Het is meer een opgave voor een rijpe onderzoeker die een beperkte bijdrage schrijft voor iets als een "Annual Review of Psychiatry/Psychology/Sociology". Hebben deze onderzoekers deze rijpheid? Gebaseerd op dit rapport moeten we deze vraag negatief beantwoorden.

IV: Derde deel: Achtergronden en meningen van daders.

Dit stuk heeft betrekking op acht plegers van geweld. Dat is te weinig. De jongens zijn niet aselekt gekozen uit een bestand van geweldplegers. Dit is ook een bedreiging voor de externe validiteit. Maar het vervelendste is dat er geen controlegroep is. Zijn deze jongeren anders dan jongeren die ook criminaliteit plegen maar geen geweld? Zijn er ook jongeren met dezelfde achtergrond die helemaal niet crimineel zijn?

Er is geen enkel meisje in de groep; het meisje dat geïnterviewd zou worden is niet op de afspraken verschenen. Dit is een mager excuus. Verder is er in dit kleine groepje ook een jongen die 3 jaar boven de leeftijdsgrens zit (20 jaar). Zoiets komt over als een slordige en onzorgvuldige benadering hetgeen gemakkelijk anders gekund had. Deze kritiekpunten ondergraven ons inziens de waarde van dit deelonderzoek. De beschrijvingen van de acht gevallen zijn interessant om te lezen, ter illustratie van het eerdere literatuuronderzoek, maar gaan niet verder dan een journalistieke benadering. Dit hoofdstuk is veel te lang en draagt niet echt bij aan de waarde van het eindproduct.

V: Vierde deel: Sleutelfiguren

De sleutelfiguren zijn zo divers dat ook dit deel niet erg nuttig is. Er zijn negen sleutelfiguren die een volstrekt verschillende achtergrond hebben. Wat heb je hier aan? Voor hetzelfde geld had je iemand uit het onderwijs met een andere mening getroffen. Deze mensen geven hun mening. Zit men daar in de Tweede Kamer op te wachten?

VI Slotbeschouwing

Het deel over mogelijke verklaringen voor de doename van jeugdgeweld (6.2) is zwak. De onderzoekers vermelden bijvoorbeeld het groeiend aantal kinderen in een-oudergezinnen als een mogelijke factor - deze opvatting weerspiegelt weliswaar een populaire opvatting, maar een rapport zoals dit zou moeten onderkennen dat het onderzoek op dit gebied minder eenduidig is als gewoonlijk verondersteld wordt. [Er is op dit gebied veel Amerikaans onderzoek]. Verwijzingen naar de invloed van de media op geweld lijken ook niet veel dieper te gaan dan *common sense* ideeën - het ligt niet zo eenvoudig als hier voorgesteld wordt. In dit gedeelte vinden wij hoegenaamd geen verwijzing naar economische of andere structurele factoren op macro-niveau welke op bepaalde groepen jeugdigen een meer criminogene invloed zullen hebben dan op andere groeperingen.

Tenslotte, de beleidsaanbevelingen (6.3). Het is een redelijke (maar niet erg originele) aanbeveling om te trachten meer complete en betrouwbare gegevens over jeugdcriminaliteit te verzamelen. Daar kan niemand bezwaar tegen maken. De beleidsaanbeveling om in verder onderzoek de sociologische benadering niet te gebruiken, maar het in dit rapport gepresenteerde levensloopperspectief te gebruiken bij de benadering van dit vraagstuk gaat ons inziens te ver en is niet onderbouwd door de inhoud van het rapport. De onderzoekers maken een te simpele tweedeling tussen sociologische en ontwikkelingspsychologische benaderingen, terwijl in de praktijk deze benaderingen elkaar juist aanvullen (zie bijvoorbeeld het werk van Sampson en Laub die de micro- en macro-levels integreren - en zo zijn er diverse andere recente voorbeelden). De beleidsaanbevelingen gericht op vroege risico-factoren sluiten aan bij de inhoud van het rapport.

VII Samenvattende waardering in woorden:

- de leesbaarheid varieert nogal; sommige delen zijn beter geschreven dan andere delen; sommige delen zijn van middelmatige kwaliteit, geven een rommelige en onnauwkeurig geformuleerde indruk;
- de relevantie voor beleid: matig; de algemene conclusie dat bepaalde risico-factoren op vroege leeftijd onderkend kunnen en moeten worden, heeft beleidsrelevantie; echter, hoe die conclusie uitgewerkt moet worden in concrete maatregelen blijft vaag. De beleidsrelevantie wordt ook ondergraven door de eenzijdige en onvolledige literatuurstudie die aan het rapport ten grondslag ligt;
- de juistheid: de interpretatie van de statistische gegevens zijn juist; het beeld van de *state of the art* van wat we weten over jeugdgeweld door middel van de literatuurstudie is onvolledig en niet altijd juist;
- de verantwoording is goed, althans voor wat betreft de statistische gegevens;
- de tijdige beschikbaarheid voor het beleid is beperkt: Het project is een response op een Tweede Kamervraag. Er ligt bijna drie jaar (30 mei 1996 Project voorstel 'gewelddriminaliteit jeugdigen' en 22 maart 1999 'aanbieding stukken bestuursraad') tussen het begin en het eind van dit project. Dat is lang. Veel langer dan nodig is voor een project dat voornamelijk bestond uit literatuurstudie en analyse van bestaande statistiek;
- de beschikbaarheid voor beleid/verspreiding: De publicatie ziet er goed uit, geschikt voor

externe verspreiding. Het is niet duidelijk aan wie deze publicatie gestuurd is, ook niet hoeveel exemplaren er gedrukt zijn;

- de inbedding in de literatuur: dit project is grotendeels een literatuurstudie. De literatuurstudie kent echter belangrijke beperkingen (zie boven);
- de inbedding in het WODC-onderzoeksprogramma: het onderwerp jeugd en geweld past goed in het WODC-onderzoeksprogramma;
- de inbedding in het justitiebeleid: onderzoek past uitstekend in justitiebeleid dat momenteel als één van de speerpunten 'jeugdgeweld' heeft.

Het is niet duidelijk welke kosten gemaakt zijn door het WODC om dit onderzoek te doen. Vanuit een kosten-baten perspectief is het uiteraard van groot belang dat in een project zoals dit onderzoekers werken die al een zekere mate van deskundigheid op het gebied van geweld en jeugdcriminaliteit hebben voordat ze aan het project beginnen. Als dit niet het geval is, zou zo'n onderzoek wellicht extern gedaan moeten worden (bijvoorbeeld door het uit te besteden aan onderzoekers die aan de universiteit verbonden zijn). We hebben de indruk dat dit onderzoek uitstekend gedaan had kunnen worden door een (1) ervaren (externe) onderzoeker met achtergrond op het gebied van statistische analyse van bestaande gegevens, en een (1) ervaren (externe) onderzoeker met goede kennis van (de theorie van) geweldscriminaliteit - en dat tegen een betrekkelijk geringe kostprijs.

Bijlage 5:

Oordeel ten aanzien van WODC 35:

EEN VERGELIJKING VAN CONFRONTATIEMETHODEN: OSLO, VIDEO EN FOTO'S

Het rapport beschrijft een experiment waarbij ten aanzien van een geënceneerde gebeurtenis aan een aantal proefpersonen (337, waarvan er 35 afvielen) gevraagd werd de "dader" aan te wijzen langs één van de volgende vijf wegen: confrontatie in persoon, simultaan (1) (Oslo-confrontatie), door het kiezen van de "dader" uit fotomateriaal, dat of simultaan of sequentieel aan deze personen werd voorgelegd (2 en 3), en door het aanwijzen van de "dader" op video, hetzij statisch, hetzij dynamisch, maar telkens sequentieel (4 en 5).

Het experiment werd op drie verschillende plaatsen (Zutphen, Amsterdam en Soesterberg) uitgevoerd. In elk van die plaatsen was de samenstelling van de groep proefpersonen wat geslacht, leeftijd en opleidingsniveau betreft gelijk. Het rapport geeft echter geen inzicht in de vraag of deze gelijksoortige groepen ook tot gelijksoortige uitkomsten kwamen.

In het rapport wordt er, zeer terecht, bij herhaling op gewezen dat de herkenning van een persoon als dader in strafzaken een zeer ernstige zaak is voor degene die als dader wordt aangewezen, zodat vergissingen en gissingen zoveel mogelijk moeten worden vermeden als het aankomt op aanwijzing van de dader. Het behoorde niet tot de taak van de onderzoekers duidelijk te maken dat elk van de confrontatie-methoden van het experiment dat plaatsvond onder gunstige waarnemingsomstandigheden, een groot aantal missers opleverde. Als de "dader" aanwezig was varieerde het aantal incorrecte aanwijzingen bij de diverse methoden van 25 tot 31 % en als de "dader" afwezig was van 43 tot 71%. Zou men die getallen kunnen generaliseren dan zal iedere advocaat die zijn vak verstaat bij elke herkenning zoveel twijfel kunnen zaaien dat op grond van deze herkenning alleen een veroordeling bijna uitgesloten is: ook 25% incorrecte herkenningen levert "reasonable doubt" ten aanzien van elke herkenning op. In verband met het noemen van percentages vermeld ik dat het toch wel nadere uitleg vergt als op pagina's 17 en 18 correcties voor toeval (bij modelvoorspellingen) zo enorm hoog zijn, bv. bij de Oslo-confrontatie moet het

percentage correcte modelvoorspellingen om die reden van 67% teruggebracht worden naar 26,5%

Het gaat in dit onderzoek om een vergelijking van de doeltreffendheid van vijf methoden ter herkenning van een dader. De slotsom is dat onder de omstandigheden van het experiment "de dader met alle methoden even vaak correct werd geïdentificeerd, zij het dat in selecties van proefpersonen waarbij de "dader" niet aanwezig was (loos alarm) de Oslo-confrontatie minder incorrecte aanwijzingen gaf dan de andere methodes. Dat dit onderzoek de resultaten van vroegere studies bevestigt (zie bijlage 6), kan ik niet helemaal onderschrijven: de onderzoekers vestigen er de aandacht op (in de samenvatting en in het rapport zelf) dat het, anders dan in vorige onderzoeken waarin consistent werd aangetoond dat bij een sequentiële foto-presentatie minder loze alarmen voorkomen dan bij een simultane, bij dit experiment niet uitmaakt of de foto-presentatie simultaan of sequentieel is. Zij proberen dat te verklaren uit verschillen (duur van de waarneming enz.) maar dit zijn niet onderzochte hypothesen.

De beschrijving van het experiment heeft bij tweede, meer indringende lezing van het rapport, aanzienlijk meer geboeid dan bij de eerste, wat oppervlakkiger kennisneming. De wetenschap gaat meestal maar voetje voor voetje, van experiment naar experiment, vooruit. De rechtspraak kan op definitieve uitkomsten niet wachten en wordt, zoals ik uit ervaring weet, vaak geconfronteerd met de vraag of iemand, de verdachte, door een waarnemer terecht is erkend als dader. Nogal vaak aanvaardt de rechtspraak herkenningen van de dader door waarnemers op wetenschappelijk niet aanvaardbare gronden. Gelukkig is meestal de herkenning niet het enige of beslissende bewijsmateriaal. Als de verdachte heeft erkend dat hij op de plaats van het misdrijf aanwezig was maar dat een ander daar het misdrijf heeft gepleegd, is zijn herkenning als dader door een *aantal* omstanders, onafhankelijk van elkaar naast zijn erkenning dat hij ter plaatse was, uiteraard veel sterker dan een simpele, enkelvoudige herkenning van een alles ontkennende verdachte. Dat die laatste herkenning in alle methodes weinig of geen bewijs oplevert, toont het rapport, ook al gaat het daar niet over, duidelijk aan. Maar om die reden zou ik dan ook menen dat één geënceneerd gebeurtenis al zijn er nog zoveel proefpersonen bij betrokken, weinig zegt. Er zullen andere gebeurtenissen waarbij de "dader" onder andere omstandigheden wordt waargenomen en waarbij de waarnemers vanuit een andere gezichtshoek of referentiekader hem (menen te) herkennen nodig zijn, om het onderzoek voor de jurist (rechter, o.m. verdachte) belangrijke betekenis te geven.

Bijlage 6:

Oordeel ten aanzien van WODC 35:

EEN VERGELIJKING VAN CONFRONTATIEMETHODEN: OSLO, VIDEO EN FOTO'S

Uit het rapport valt niet duidelijk op te maken wie er in de begeleidingscommissie zitting hadden. Dat is jammer. Het is uitgevoerd door onderzoekers van TNO (Technische menskunde), samen met de Rechterschool in Zutphen.

Het huidige onderzoek is een voortzetting en uitbreiding van eerder soortgelijk onderzoek naar de effectiviteit van vijf verschillende dader identificatietechnieken: video, Oslo en foto door dezelfde onderzoekers. De probleemstelling en vraagstelling vloeien rechtstreeks voort uit het eerdere onderzoek. Echter, hetgeen benadrukt wordt in het document 'aanbieding stukken bestuursraad' (OE-382/17.98) dat bij dit rapport hoort, vindt men in de 'vraagstelling experiment' (p.10) niet terug. Tenminste, de nadruk ligt heel anders: "De *veranderingen in instructie en aanbiedingstijd* ten opzichte van het vorige onderzoek hadden op zichzelf geen effect op het

aantal correcte en niet-correcte identificaties. Gezamenlijk hadden die twee veranderingen dat (effect) wel..." staat er in het document 'aanbieding stukken bestuursraad'. Dus, het document 'aanbieding stukken bestuursraad' presenteert als centrale vraagstelling de vergelijking tussen het eerste en het tweede onderzoek, terwijl het rapport zelf dit maar terloops doet (zie vraagstelling experiment, p.10).

Tenslotte, een andere - kleine - opmerking betreft de 'vraagstelling experiment' (p. 10) waar onder punt drie wel de simultane en de sequentiele presentatie van personen op de foto expliciet genoemd worden, maar eenzelfde vergelijking wat betreft de videopresentatie niet vermeld wordt.

Het is bij uitstek een pragmatisch, beleidsgericht onderzoek, wat ook blijkt uit de betrekkelijk oppervlakkige bespreking van relevant onderzoek en theoretische kwesties. Die is nogal magertjes. Het aangehaalde onderzoek is beperkt tot datgene wat rechtstreeks betrekking heeft op het empirisch onderzoeken van de betrouwbaarheid van ooggetuigen-verklaringen, zonder aandacht te schenken aan de zeer rijke en uitgebreide kennis die bestaat in de cognitieve psychologische literatuur over geheugen, *memory decay*, *memory retrieval*. Dit zou tot meer genuanceerde hypothesen hebben kunnen leiden. De onderzoeksvraag vertegenwoordigt echter een aanvaarde benadering binnen het veld, en - gezien het doel van het onderzoek - is een uitgebreider literatuuronderzoek wellicht onnodig.

De keuze van het design (het klassieke experiment) is goed. Immers, de onderzoeksvraag is bij uitstek geschikt voor een *experiment*, waar onder constant gehouden condities, één variabele (de confrontatiemethode) op vijf verschillende manieren geoperationaliseerd wordt (statische confrontatie in persoon - Oslo -, videoconfrontatie - zowel statisch als simultaan - en fotoconfrontatie - zowel statisch als simultaan). De resultaten worden geëvalueerd op het criterium *effectiviteit* (d.w.z. het aantal correcte identificaties en het aantal incorrecte identificaties, de zogenaamde 'valse-positieven') van de vijf verschillende methodes. Interessant is dat de onderzoekers - behalve effectiviteit - ook *kosten* en de *praktische uitvoerbaarheid* in de discussie (maar niet als onderdeel van het experiment) betrekken. Er wordt door de onderzoekers - waarschijnlijk terecht - *verondersteld* dat de foto-methode en de video-methode sowieso goedkoper en gemakkelijker uitvoerbaar zijn dan de Oslo-methode. Opvallend is dat in Tabel II, waar de voor- en nadelen van confrontatie in persoon, fotoconfrontatie, en de videoconfrontatie op een rijtje gezet worden, geen vermelding gemaakt wordt van de lagere kosten. Dus, het feit dat (hun eigen) eerdere onderzoek aangetoond heeft dat de Oslo-methode wat het criterium effectiviteit betreft het beste is (m.n. door het minder voorkomen van incorrecte identificaties), moet in het rapport verzoend worden met de realiteit dat de praktijk waarschijnlijk de voorkeur geeft aan video's en foto's. Daar is niets op tegen, mits het in het rapport expliciet gemaakt wordt wat de verschillende overwegingen zijn. Dat gebeurt ook.

De beschrijving van de methode van het experiment is goed te volgen. De beslissing om als proefpersonen alleen mensen te werven via (muziek) verenigingen is enigszins curieus, of behoeft op zijn minst wat meer uitleg. Bijvoorbeeld, komen de mensen alleen uit muziekverenigingen, of ook uit andere verenigingen? Zo ja, wat voor verenigingen? Hoe werden de mensen geselecteerd en gerecrueteerd? Waren er veel weigeringen? Het is waar dat de representativiteit van de steekproef niet van doorslaggevend belang is om de resultaten van experimenten naar de gehele populatie te kunnen generaliseren, maar toch lijkt de huidige steekproef wat zwak op het punt van de externe validiteit. Zo behoren alleen een bepaald soort mensen tot een (muziek)vereniging. Op zijn minst zouden de onderzoekers meer informatie kunnen verschaffen over de sociaal-demografische kenmerken van de steekproef (ter vergelijking met de gehele populatie).

De beschrijving van de procedure is duidelijk. Een kleine opmerking betreft het niet vermelden van het aantal proefpersonen dat geen toestemming gaf om hun gegevens te gebruiken (p.13). De

logic van het experiment is helder uiteengezet.

Wat de *resultaten* betreft, het volgende. Bij de beschrijving van de proefpersonen wordt vermeld dat er een indeling is gemaakt in vier categorieën, waaronder lagere school, met slechts drie (3) personen in deze categorie. Dat is veel te weinig, deze categorie zou samengevoegd moeten worden met lagere opleiding.

De resultaten zijn in begrijpelijke vorm (tabellen) gepresenteerd en duidelijk samengevat. Interessant is dat dit experiment de resultaten van het eerdere experiment (waarop de huidige studie gebaseerd is) bevestigt. Enigszins onduidelijk is het volgende: Op pagina 21 staat: "In voorgaande studies werden proefpersonen veelal geïnstrueerd om na elke presentatie van een persoon aan te geven of die persoon de 'dader' was of niet. In het huidige experiment werd aan proefpersonen medegedeeld dat zij direct moesten melden als de bedoelde persoon zich in de selectie bevond..." Ik zie geen verschil in de twee methoden - tenminste zoals het hier beschreven wordt?

Tenslotte zou het interessant geweest zijn om het belang van de 'cues' nader te becommentarieren. Op pagina 22 wordt gesteld dat (1) fotoconfrontaties minder 'cues' geven voor herkenning -waardoor vaker de 'incorrecte' figurant geselecteerd wordt, en dat (2) als men langer kijkt, de kans groeit dat het oordeel wordt gebaseerd op minder discriminerende 'cues' (waardoor men vaker incorrect raadt). Hoe zijn die observaties verenigbaar?

Dit is een goed rapport, met mogelijk bruikbare beleidsimplicaties. Het experiment lijkt zorgvuldig uitgevoerd te zijn, volgens aanvaardbare methodes. Tenslotte, de aanbevelingen voor verder onderzoek zijn theoretisch weinig onderbouwd - zonder theoretische argumenten ter ondersteuning van verdere experimenten kan men nog vele jaren doorgaan met dit soort experimenten, door iedere keer een kleine variant van het oorspronkelijke *design* te introduceren. Bijlage 7

Oordeel ten aanzien van WODC 16:

VERGUNNINGVERLENING BIJ BEDRIJFSVESTIGING

In het kader van de MDW-operatie wordt aan bureau Oranjewoud gevraagd om vanuit de optiek van de bedrijven te inventariseren hoe er wordt geoordeeld over de milieuvergunningenprocedure. Het onderzoek heeft een sterk beschrijvend en inventariserend karakter. De onderzoekers onthouden zich van aanbevelingen en laten dit over aan de ambtelijke stuurgroep. In de begeleidingscommissie van het onderzoek is het WODC niet vertegenwoordigd. Er is geen onderzoeksdeskundigheid ingebracht. Het beleidsveld is zeer selectief vertegenwoordigd. Er is hier sprake van een nogal besloten 'poldermodel'. Een opvallend kenmerk van het onderzoek is het ontbreken van literatuurverwijzingen. Er is reeds veel onderzoek gedaan naar de optiek van bedrijven op processen van vergunningverlening (ECW/EIM/CSTM), maar daar wordt niet naar verwezen.

De optiek van het onderzoek is beperkt. De procedures worden niet beoordeeld naar de mate waarin ze bijdragen aan een inhoudelijk (juridisch/milieukundig/planologisch) goed besluit (zie doel van de betreffende wetgeving). Er wordt vanuit een beperkt doelmatigheidsaspect (beperkt houden van transactiekosten) naar deze procedures gekeken. Het soort aanbevelingen dat wordt verwacht, is eveneens van te voren gegeven. Tijdsdruk leidt tot een zogeheten quick scan.

Het is gelet op het op processen-gerichte karakter van het onderzoek opmerkelijk dat er geen vragen worden gesteld aan rechters, rechtshulpverleners, bezwaarmakers (milieu-organisaties). Aan het onderzoek gaat geen juridische analyse van het geldende recht vooraf.

Grote delen van het onderzoek (deel 1) bevatten een eerste verslaglegging van de gegevensverzameling. De leesbaarheid wordt daardoor negatief beïnvloed.

Aangezien het onderzoek mikt op representatieve uitspraken is de selectie van de cases in Friesland en Groningen opmerkelijk. Onduidelijk is waarom het onderzoek tot deze provincies beperkt blijft. Kennelijk spelen twee doelen door elkaar heen. De provincies Friesland en Groningen die zowel kampen met enige werkloosheid als met veel te veel bouwrijp gemaakt bedrijventerrein hebben in het recente verleden Oranjewoud beeldvormingsbeïnvloedend onderzoek laten doen om aan te tonen dat de Waddenzee-autoriteiten in Noord-Nederland echt niet zwaarder toetsen dan het bevoegd gezag in de andere gebieden langs zeearmen (Westerschelde) en de grote rivieren (Maas-Waal). De opzet van dat onderzoek (inclusief case-selectie) wordt nu (blijkbaar wegens doorslaand succes) geprolongeerd.

De case-selectie is kwetsbaar. De noordelijke provincies hebben slechts een beperkt deel van de industriële werkgelegenheid. De bestuursorganen stellen zich bovenmatig bedrijfsvestigingsvriendelijk op. Er is geen sprake van een bedrijfsterreinentekort. Er wordt niet geselecteerd. Dit geheel in tegenstelling tot Rijnmond, Brabant en Randstad (in welk gebied de 'bulk' van de vergunningplichtige bedrijven kan worden aangetroffen). Kortom niet erg 'representatief'.

De selectie van Mijnewet-vergunningen is bepaald opmerkelijk. Komt immers weinig voor. Eerder de WVO-vergunning. De beschrijvende gegevens in de eerste schema's van hoofdstuk 9 zijn zeker niet 'representatief'. Alternatieve, representatieve beschrijvingen hadden zonder veel moeite verkregen kunnen worden via de statistische systemen van VROM (milieuinspectie), via een enquête bij de provincies, of via de ECW.

Het is opvallend dat het verzamelde case-study-materiaal wel met elkaar wordt vergeleken, zodat er een soort gemeenschappelijke noemer ontstaat, maar er vindt geen (op verschillen gerichte) analyse plaats waardoor bepaalde samenhangen tussen variabelen ontdekt zouden kunnen worden.

Dat het project 'theorie-arm' en 'beschrijvend-beoordelend' is, levert risico's op bij het doen van aanbevelingen tot aanpassing van het huidige stelsel. Deze aanbevelingen zijn dan wellicht gebaseerd op gemeenschappelijke knelpunten, maar de onderzoekers kunnen van te voren niet aangeven wat de (mogelijke) effecten van de doorvoering van deze aanbevelingen zouden kunnen zijn.

De aanbevelingen van de ambtelijke stuurgroep op basis van dit onderzoek zijn zeer algemeen. Zijn niet nieuw. Zijn niet erg onderbouwd. Verdienen nadere uitwerking. Wat dat betreft zou het echte onderzoek nog moeten beginnen.

Bijlage 8:

Oordeel ten aanzien van WODC 12:

EVALUATIE POLITIEWET 1993: BREEDTE-ONDERZOEK

De breedtestudie is betrekkelijk snel uitgevoerd (korte doorlooptijd, snelle tussenrapportage). Het gaat om de beschrijving van de formele structuur en het feitelijke functioneren van de 25 politieregio's. Het eerste deel van het onderzoek (hoe is formele structuur) heeft plaatsgevonden door het opvragen van documenten (en het later analyseren en rapporteren daarover). Het tweede deel van het onderzoek (hoe functioneert het feitelijk, zijn er knelpunten) heeft plaatsgevonden door het houden van vijf telefonische interviews met opinionleaders (experts/belanghebbenden) in de 25 politieregio's.

Het doel van de beide breedteinventarisaties is tweeledig. Deels wordt geprobeerd om verschillen tussen de politieregio's op het spoor te komen ten behoeve van de case-selectie voor het daarop aansluitende kwalitatieve diepteonderzoek. Daarnaast worden ook in de analyserende hoofdstukken typologieën ontwikkeld om de verschillen tussen de politieregio's te vangen, stijlen van officieren te typeren, stijlen van korpsbeheerders en korpschefs te typeren. Kortom er wordt een overzicht geboden.

Het andere doel van het onderzoek lijkt te zijn om een algemeen, gemiddeld beeld (de grootste gemene deler) te geven van alle 25 regio's zodat een oordeel gegeven kan worden over de werking van de Politiewet (uitgesplitst naar democratische inbedding, daadkracht en dualisme). Positief aan het onderzoek is de bijlage 1 met de regioprofielen van de 25 politieregio's. Dit is in feite een beschrijvende datamatrix. Het stelt de lezer in staat om zelf de diversiteit van de wetsuitvoering (van hoofdstuk 5 Politiewet) te onderkennen.

Kritische kanttekeningen bij het onderzoek zijn:

- de constatering ten aanzien van de democratische inbedding (hoofdstuk 5) bevatten weinig meerwaarde boven hetgeen al uit een bestudering van de wetstekst opgemaakt had kunnen worden;
- de constatering over bestuurlijke daadkracht (hoofdstuk 4) beloven meer dan er wordt waargemaakt. Feitelijke beschrijving en algemeen onderschreven knelpunten zijn niet hetzelfde als de vaststelling dat er sprake is van gebrekkige bestuurlijke daadkracht. Het centrale begrip bestuurlijke daadkracht wordt eigenlijk niet uitgewerkt. Het blijft een voor empirisch onderzoek onhanteerbaar begrip.
- ook het begrip 'dualisme in beheer' (hoofdstuk 6) heeft als centraal begrip weinig (operationele) uitwerking gekregen.

Per saldo is het de vraag of er wel sprake is geweest van een evaluatie in de eigenlijke (meer strikte) zin van het woord. De situatie voor invoering van de Politiewet wordt niet in kaart gebracht. De gewenste situatie van de makers van de Politiewet wordt niet erg scherp gedefinieerd als mogelijke beoordelingsmaatstaf. Wel worden de globale doelen van de wet genoemd (maar niet uitgewerkt tot scherpe maatstaven). De huidige feitelijke situatie wordt wel beschreven. Deels op basis van formele documenten, deels op basis van oordelen van betrokkenen. In hoeverre die feitelijke situatie mede het resultaat is van veranderingen die het gevolg zijn van de invoering van de Politiewet blijft onbeantwoord. Bij de beoordeling van de vraag of de wetwijziging de gewenste doelen heeft bereikt, wordt door de onderzoekers volstaan met de weergave van de knelpunten zoals deze door de direct-betrokkenen (deels ook belanghebbenden) in de gesprekken worden genoemd. Kortom het is een aardig beschrijvend (inventariserend) (knelpunten)onderzoek, maar het voldoet niet aan de meer stringente eisen van een effectiviteitsonderzoek (de causale toerekening van veranderingen aan de invoering van de wet; en de beoordeling van de netto-effecten aan de beoordelingsnormen van de wetgever vinden immers allebei niet plaats). Wel is de werkwijze van de onderzoekers goed gedocumenteerd. Goed valt te controleren hoe de onderzoekers tewerk zijn gegaan.

We weten inmiddels dat dit onderzoek meer dan 300.000,- gulden heeft gekost. Men zou kunnen zeggen dat het wel wat duur is uitgevallen. Positief lijkt mij te zijn dat het er betrekkelijk tijdig lag. Het 'eindproduct' is op tijd opgeleverd.

Bijlage 9:

Oordeel ten aanzien van WODC 13:

EVALUATIE POLITIEWET 1993: DIEPTE-ONDERZOEK

Na een openbare EU-aanbestedingsprocedure werd een consortium gecontracteerd. Het betreffende onderzoeksverband omvat drie vakgroepen, van drie universiteiten. Er zijn vijf redacteurs van het eindrapport, er zitten drie leden in de interne 'reflectiegroep', tien onderzoekers hebben medewerking gegeven aan de dataverzameling/rapportage. Het lijkt er op alsof de complexe bestuurlijke organisatie van hoofdstuk 5 van de Politiewet (het object van het onderzoek) zich heeft voortgeplant in de complexe organisatie van het onderzoek zelf.

Er is een externe begeleidingscommissie (op enige afstand van het WODC) met een kennelijk groot mandaat. Er wordt door deze externe begeleidingscommissie een dubbelslag gemaakt door een breedte- en een diepte-onderzoek te onderscheiden. Het diepte-onderzoek is voorwerp van openbare aanbesteding.

Men kan zich de vraag stellen of hier niet te zware organisatorische maatregelen op een op zich betrekkelijk eenvoudige vraag zijn gezet. In het kwalitatieve diepte-onderzoek worden zes organisatiebeschrijvingen gegeven, voornamelijk door de relevante actoren en centrale relaties in kaart te brengen. In geen van de zes cases lukt het om de organisatiestructuur in relatie te brengen tot de productie (activiteiten; prestaties) van de zes regiopolitiekorpsen.

Het voordeel van het diepteonderzoek boven het breedteonderzoek lijkt te zijn dat er iets meer aandacht gegeven kan worden aan de tijdsdimensie. De verschillen tussen de eerste (formatieve) jaren van de politieregio en de latere meer (ervaren, rijpere) jaren ten tijde van het onderzoek komen in het diepte-onderzoek goed uit de verf.

Een andere toegevoegde waarde van de dieptestudie boven de breedtestudie is dat de lagere managementniveau's van de politieorganisatie (district, lokaal, dagelijks bestuur, presidium) scherper in beeld zijn gekomen. Tenslotte komen de verschillen tussen de zes cases in de dieptestudie door de veelheid van informatie veel duidelijker naar voren dan de verschillen tussen de 25, meer oppervlakkig beschreven, politieregio's in de breedtestudie.

In de dieptestudie wordt aan het eind een normatief oordeel gegeven over het huidige feitelijke functioneren van de politieregio. Bovendien worden er ook aanbevelingen gedaan over de adequaatheid van de huidige Politiewet. Deze aanbevelingen berusten op beschrijvende informatie uit de casushoofdstukken en zijn ook te herleiden tot algemene uitspraken over de doelen van de wetgever, maar berusten niet op een uitgewerkt normenkader en ook niet op een uitgewerkte (causale) analyse van de verschillen tussen de zes cases.

Het rapport eindigt met een honderdtal uitspraken die vaak een hoge beleidsrelevantie hebben en vaak ook redelijk goed onderbouwd (beredeneerd) kunnen worden met het verzamelde materiaal.

- De beschrijving van de zes cases is van wisselende kwaliteit. Men kan merken dat verschillende onderzoekers aan het werk zijn geweest, hoewel de structuur van de caseverslagen wel in hoge mate dezelfde is (er is met een strak keurslijf gewerkt, maar niet alle teams konden dat blijkbaar even goed hanteren).

Oorspronkelijk zijn de zes cases gekozen omdat ze verschillende posities innemen op de dimensies centralisatie/decentralisatie en formalisatie/informeel. Deze verschillen blijken er ook te zijn, maar er wordt niet afzonderlijk geanalyseerd waardoor deze structuurverschillen verklaard worden (stedelijkheid/platteland, aantal deelnemende gemeenten, omvang korps als geheel), noch

wordt een poging gedaan deze verschillen in bestuurlijke structuur te verbinden met verschillen in knelpunten, managementstijlen en werkwijze.

Per saldo is het diepteonderzoek erg duur uitgevallen. Zes en een halve ton. De besluitvormingsprocedure om tot deze opdrachtnemeselectie te komen lijkt nogal zwaar geweest te zijn. Negen instituten hebben offerte uitgebracht. De selectiekosten (transactiekosten) van deze wijze van aanbesteding zijn hoog geweest. Het is opvallend dat de kennelijk (met de EU-richtlijn) beoogde marktwerking niet echt is gerealiseerd.

Opvallend is dat er geen case-vergelijking plaatsvindt tussen de zes politieregio's, dat de hypothesen/typologieën van het ITS niet worden 'getoetst'. Wat dat betreft is de afstemming tussen breedte- en diepte-onderzoek niet optimaal. Positief is natuurlijk dat dit omvangrijke (en belangrijke) onderzoek op tijd af was. Dat pleit voor het organiserend vermogen van de drie onderzoeksteams.

Het onderzoek wordt het 'diepteonderzoek' genoemd. Dit roept beelden op van gedetailleerde en levendige beschrijving. Dat is echter feitelijk niet het geval. De wijze van beschrijving is gestandaardiseerd, afstandelijk, globaal en heeft weinig oog voor de details. Er is geen duidelijke (positieve) keuze gemaakt voor een meer kwalitatieve, procesmatige gevalsanalyse.

Aan de gehele 'procesarchitectuur' van de beide onderzoeken kleven wellicht gebreken. Positief zijn de complete inventarisaties van het ITS en de normatieve interpretaties van de gegevens door de politiedeskundigen van het consortium (het diepteonderzoek), plus het feit dat er tijdig is geleverd en dat er tussen de teams geen noemenswaardige spanningen/conflicten/verschillen van inzicht zijn opgetreden (er was sterke procesbeheersing).

Negatief in het gehele onderzoekscircus zijn: de kosten (meer dan een miljoen), de tijdsdruk (te laat begonnen, waarom niet eerder een eigen initiatief van het WODC, waarom wachten tot de politieministers de eigen interne evaluatie Politiewet in 1998 hadden uitgevoerd?), de hoge transactiekosten (veel overleg en georganiseerd), het ondergeanalyseerd zijn van het gepresenteerde materiaal, de globaal blijvende cases in het diepterapport en de niet ver uitgewerkte typologieën in het breedterapport. Kortom, de (binnengehaalde, universitaire) onderzoekswereld betaalt wel een prijs voor deze beleidsmatige tijdsdruk. Er wordt te weinig gereflecteerd op het vele materiaal dat werd verzameld (ook dat is een vorm van verkwisting).