

Startnotitie WODC-onderzoek

Aan

Van afdelingshoofd

Concipiënt Frank Willemsen

Doorkiesnummer(s) 6847

Datum 15 januari '09

Projectnummer 1761-2

Werktitel van het project

PROGRAMMERINGSTUDIE: DEEL II DESIGNS EFFECTMETINGEN HANDHAVING EN TOEZICHT

Doelstelling van het project

Het doel van dit onderzoek is het zoeken naar en presenteren van onderzoeken die op valide wijze effecten van toezicht (of van met toezicht vergelijkbare activiteiten) meten. De zoektocht wordt breed ingezet zodat ook onderzoeken buiten het toezicht- en handhavingsdomein worden meegenomen. De resultaten moeten bijdragen aan de inrichting van twee onderzoeksprogrammeringen (handhaven & gedrag en effectief toezicht).

Probleemstelling

- Welke intern valide effectonderzoeken gaan over handhaving en toezicht, welke designs zijn gebruikt en wat zijn de uitkomsten?
- Welke intern valide effectonderzoeken hebben een raakvlak met de wereld van handhaving en toezicht?
- Welke andere (minder intern valide) onderzoeken hebben een raakvlak met de wereld van handhaving en toezicht en dragen bij aan kennis over effectiviteit?
- Hoe zouden deze onderzoeken kunnen worden uitgevoerd door toezichthouders en handhavers?

Beleidscontext

Inleiding

Bij verschillende overheidsorganisaties leeft de behoefte aan meer wetenschappelijke kennis over toezicht en handhaving en de naleving van regelgeving. Ook in de (internationale) wetenschappelijke literatuur over handhaving en de (effecten van) handhaving en toezicht wordt geconstateerd dat er (te) weinig valide en betrouwbaar onderzoek wordt gedaan naar handhaving en toezicht en de effectiviteit ervan.

Om het *gedragwetenschappelijk onderzoek* op dit domein te bevorderen, is in 2005 het interdepartementale onderzoeksproject *Handhaving en gedrag* gestart¹. Het project is gericht op het stimuleren van het vergaren van empirische wetenschappelijke kennis met een gedragswetenschappelijk karakter over de naleving en de handhaving van zgn. ordeningswetgeving. In 2008 is besloten tot een professionalisering van het onderzoeksprogramma door op basis van een uit te voeren programmeringstudie een Meerjarenprogramma Handhaving en Gedrag op te stellen.

Binnen de Inspectieraad leeft ook al enkele jaren de ambitie om meer inzicht te verkrijgen in de effecten van toezicht. Effectmeting staat bij alle rijksinspecties evenwel nog in de kinderschoenen. Om een impuls te geven aan het effectmetingsonderzoek heeft de Inspectieraad in 2008 de werkgroep 'Evaluatie effecten van toezicht' opgericht. Doel van het programma van de werkgroep 'Evaluatie effecten van toezicht' is de toezichtspraktijk te verbeteren door gebruik

¹ *Handhaving en gedrag* is een samenwerkingsproject van de Belastingdienst, de Inspectie Gezondheidszorg, de Inspectie Verkeer en Waterstaat, de Voedsel en Watenautoriteit, de VROM-inspectie, gecoördineerd door het Expertsiecentrum Rechtspleging en Rechtshandhaving van het ministerie van Justitie.

te maken van resultaten van (wetenschappelijk) onderzoek van de toezichtspraktijk en de effecten ervan. Hiertoe zal onder meer een meerjarenplanning worden opgesteld voor de Inspectieraad.

Deze startnotitie is geschreven t.b.v. de aanbesteding van de programmeringstudies en dient uiteindelijk bij te dragen aan de opzet van een (wetenschappelijk) onderzoeksprogramma voor het *meerjarenprogramma Handhaven en Gedrag* en het *meerjarenprogramma Evaluatie effecten van toezicht*. Het doel van de opdracht is het in beeld brengen van de beschikbare (wetenschappelijk) kennis over de relatie tussen handhaven en toezicht, gedrag en de effecten. Deze startnotitie gaat over *designs* en *best practices* die gebruikt kunnen worden om effecten van toezicht en handhaving aan te tonen. Een ander deel van deze studie gaat over gedragsmechanismen die ten grondslag liggen aan handhaving en toezicht en wordt in een separate startnotitie behandeld. Tenslotte is er een derde startnotitie die over de synthese van de onderzoeksresultaten uit beide studies gaat en aandacht besteedt aan de totstandkoming van twee onderzoeksprogrammeringen (handhaven & gedrag en effecten van toezicht).

Achtergrond

Een steeds terugkerend thema in (internationale) wetenschappelijke literatuur over handhaving en toezicht is de constatering dat er (te) weinig valide en betrouwbaar onderzoek wordt gedaan naar de effectiviteit hiervan.

Onder *toezicht* wordt verstaan het verzamelen van informatie over de vraag of organisaties, beleid of voorzieningen aan de gestelde eisen voldoen, het zich daarna vormen van een oordeel daarover en het eventueel naar aanleiding daarvan interveniëren.

De Commissie Handhaven op Niveau verstaat onder *handhaving* de zorg voor de naleving van regels. Handhaving omvat in deze visie alle activiteiten om de naleving van regelgeving te bevorderen: zowel preventieve activiteiten (bijvoorbeeld voorlichting over de regels) als repressieve activiteiten (sancties).

Organisaties die toezicht uitoefenen, heten doorgaans inspectie (Inspectie Werk en Inkomen), Autoriteit (Autoriteit Financiële Markten) of Commissariaat (van de Media). Onderscheid wordt wel gemaakt tussen kwaliteitsmeesters (zoals de Onderwijsinspectie), marktmeesters (zoals de Opta en de NMA) en regelmeesters (zoals de Arbeidsinspectie). Ook verschillende objecten van toezicht worden (o.a. door de Ambtelijke Commissie Toezicht) onderverdeeld in:

- Toezicht op zelfstandige organisaties en professionals met een publieke taak (scholen, ziekenhuizen, woningcorporaties maar ook de leraren en artsen);
- Toezicht op een publieke of een beleidssector (onderwijs, volkshuisvesting);
- Toezicht op marktsectoren (eerlijke concurrentie, marktmeesterschap);
- Toezicht op de naleving van aspectwet- en regelgeving (milieu, arbeidsomstandigheden)

Meestal echter worden nalevingstoezicht, uitvoeringstoezicht, interbestuurlijk en markttoezicht als toezichtsvormen onderscheiden. Als toezichthouders dwang uitoefenen, wordt van handhaving gesproken (Commissie Kuipers, 2005, 10).

Een taak van een handhaver / toezichthouder is te zorgen dat wetten en regels worden nageleefd door burgers, bedrijven en de overheid zelf. Controleren van naleving is geen op zichzelf staand doel maar dient bij te dragen aan het voorkomen en beheersen van collectieve risico's die bepaalde gedragingen van onder toezichtgestelden met zich meebrengen en dient (positieve) welvaartseffecten te helpen genereren (betere gezondheid, betere opleidingen, veiliger verkeer en vervoer enz.). Voor het voorkomen / beperken van risico's hebben handhavers en toezichthouders gedragscontrolerende, gedragsmodificerende en gedragsregulerende activiteiten in huis. Er kan een globaal onderscheid gemaakt worden tussen milde en meer strikte stijlen van handhaving en toezicht, daarnaast komen mengvormen voor. De keuze voor een bepaalde stijl wordt onder andere bepaald door de ernst van het risico dat moet worden beperkt, de mate van naleving van regels en wetten, de cultuur van een toezichthouder en de verwachte effectiviteit van een bepaalde stijl.

Interventies

Toezichthouders, inspecteurs en handhavers zijn in de afgelopen 10-15 jaar komen te beschikken over meer en meer gedifferentieerde interventies. Hun interventiemacht is daarmee duidelijk toegenomen (Willemsen, Leeuw & Leeuw, 2008). Naast waarschuwingen en het intrekken van vergunningen is in circa 60 wetten voorzien dat bestuurlijke boetes opgelegd kunnen worden (Michiels, 2006) die soms kunnen oplopen tot vele miljoenen euro's. Ook worden reparatoire sancties toegepast, zoals de 'last onder dwangsom' waarmee de toezichthouder beoogt een halt toe te roepen aan een overtreding en erop uit is om, waar mogelijk, reparatie te laten plaatsvinden of de schade te laten herstellen. Bestuursdwang functioneert als stok achter de deur of voor gevallen waarin geen overtreders bekend zijn. Daarnaast is er de strafbeschikking van de politie en de OM-afdoening (de strafbeschikking in handen van het OM) (vgl. Elsink, 2007, 20-22). Tenslotte wordt met naming and shaming gewerkt, waarbij organisaties met naam en toenaam aan de schandpaal genageld worden.

Toezichthouders kunnen naast het heffen van bovengenoemde sancties ook de rol van adviseur aannemen, waarbij hij de onder toezichtgestelde met raad en daad bijstaat en er met zijn adviezen toe bijdraagt dat deze aan de gestelde eisen voldoet (WRR, 2004). Compliance Assistance (het bevorderen van vrijwillige naleving) is dan ook gebaseerd op *vertrouwen*. Door het verstrekken van informatie en het geven van ondersteuning aan de gereguleerden poogt de handhaver en toezichthouder deze te helpen met het begrijpen en naleven van regels. Systeemtoezicht is ook gebaseerd op vertrouwen maar dan in reeds bestaande, goed functionerende controlesystemen. Bij bewezen goede kwaliteitssystemen en correct naleefgedrag worden onder toezichtgestelden beloond door een groter vertrouwen van de toezichthouder en waarschijnlijk minder controles (proportioneel toezicht).

Governance (of deugdelijk bestuur) is een principe dat vooraf kan gaan aan het invoeren van systeemtoezicht. Verondersteld wordt dat wanneer organisaties werken volgens principes van (good) governance ten eerste 'compliance' met bestaande regels beter vorm krijgt, tegelijkertijd toezicht, inspectie en 'controles' wat minder op de voorgrond hoeven te treden en in plaats daarvan vooral 'horizontale' principes van het afleggen van verantwoording gehanteerd kunnen worden. Ook zelfregulerend vermogen, kwaliteitssystemen, certificering en accreditatie kunnen in verband worden gebracht met toezicht op basis van vertrouwen.

Wantrouwen versus vertrouwen

In de publieke discussie over handhaving en toezicht in de afgelopen jaren doet zich een paradoxaal verschijnsel voor. Enerzijds is er de tendens van politiek en samenleving om na incidenten en crises op te roepen tot méer handhaving en toezicht (en regels). In de Brief van de Minister van BZK over het programma 'Vernieuwing Toezicht' (Min BZK, 31 201 25) heet het zo: 'na incidenten vragen politiek en samenleving doorgaans om versterking van het toezicht'. Anderzijds roepen burgers en onder toezicht gestelde organisaties tegelijkertijd dat ze te veel last hebben van bemoeienissen van toezichthouders, door de 'druk(te)' die ze veroorzaken, de administratieve lasten die daarmee gepaard gaan en omdat lang niet altijd duidelijk is hoe toezicht bijdraagt aan het gemeenschappelijke belang.

Toezicht moet kortom anders, efficiënter en effectiever en hieraan gerelateerd gaat het Kabinet er van uit dat inspecties zoveel mogelijk gebruik gaan maken van controlesystemen die in de onder toezicht staande organisaties en ketens zelf al bestaan. (vgl. de Brief van de Minister van BZK over het programma 'Vernieuwing Toezicht' (Min BZK, 31 201 25)). Er wordt met andere woorden gepleit voor (en nogal wat verwacht van) *trust-and-hope* regulation, niet alleen in Nederland maar ook in andere westerse landen (Shover, 2008). Zelfregulering en systeemtoezicht vallen hieronder en zijn goeddeels gebaseerd op vertrouwen in onder toezichtgestelden.

De orthogonaal van "trust-and-hope regulation" is formalistisch toezicht dat ook wel als "command and control" of "going by the book" wordt getypeerd (Erp

2008). Deze vorm van toezicht is gebaseerd op wantrouwen en afschrikking en leunt sterk op bestraffing. Uiteraard vinden vertrouwen en wantrouwen en trust-and-hope en command and control elkaar in een mix. Responsive regulation (Braithwaite, 1992) is de meest bekende strategie die gebruik maakt van beide uitersten. Responsive regulation gaat uit van de handhavingspiramide die aansluit bij de gedachte "niet goedschiks dan maar kwaadschiks" en is in de afgelopen jaren uitgegroeid tot de dominante strategie van Nederlandse handhavingsorganisaties (Van de Bunt, 2007).

Designs om effectiviteit te bepalen

In deze startnotitie hebben we verschillende toezicht- en handhavingsstijlen besproken en zijn we ingegaan op de sanctierijkdom van de Nederlandse toezichthouders. We zijn echter geïnteresseerd in hoeverre elementen als stijl- en sanctietype zorgen dat onder toezichtgestelden zich (beter) aan de regels houden of nog breder gesteld, wat het maatschappelijke effect van de toezicht- en handhavingsactiviteiten is.

Wie de vraag naar de (maatschappelijke) effecten van handhaving en toezicht stelt, stelt in directe of in afgeleide zin, een causaliteitsvraag. In welke mate verscherpt toezicht op onder de maat presterende onderwijsinstellingen leidt tot (of bijdraagt aan) verbetering van de onderwijskwaliteit, in welke mate het scherper 'on site' controleren van vliegtuigen leidt tot (of bijdraagt aan) een veiliger luchtvaart of in welke mate bestuurlijke boetes zorgen voor minder recidiverende bedrijven. Om dit type vragen te kunnen beantwoorden is kennis nodig over de counterfactual: 'wat zou er gebeurd zijn (met die scholen en vliegtuigen, bedrijven) als er geen (verscherpt) toezicht was geweest of als er geen boetes waren uitgeschreven?' Op twee manieren kan de counterfactual gekend worden. De eerste manier is door gebruik te maken van redeneringen of theorieën die 'aangeven' 'what would have happened anywhere'. De tweede manier is door een evaluatieonderzoek zo in te richten dat de counterfactual 'gesimuleerd wordt'. Dat kan via het (quasi-)experimentele design, omdat een goed gekozen controlegroep handelt 'alsof er geen beleids- of inspectie-interventie is geweest'.

Het belangrijkste voordeel van een (quasi) experimenteel design is de hoge mate van interne validiteit. In de internationale literatuur over evidence-based policymaking en what works worden effectonderzoeken wel geschaald op een vijfpuntsschaal, die bekend staat als de Maryland Scientific Methods Scale (MSMS). Hoe hoger de MSMS-score, des te adequater het design en de interne validiteit van het onderzoek en daarmee de waarschijnlijkheid dat een gevonden verschil in uitkomsten aan de onderzochte interventie kan worden toegeschreven. De MSMS kent vijf niveaus die we hieronder kort zullen beschrijven:



Uitgebreid wordt de MSMS gebruikt in evaluaties van de effecten van justitiële maatregelen (Welsh & Farrington, 2006), onderwijs (Cook, 2003) en verschillende

andere beleidsvelden (vgl. Campbellcollaboration.org). Overigens wijzen Pawson en Tilley (1997) erop dat, om te begrijpen waarom iets een effect heeft of niet, aandacht voor de aan de interventie ten grondslag liggende 'theorie', wezenlijk is. Daarom besteden we in de andere startnotitie aandacht aan gedragsmechanismen die geactiveerd (dienen te) worden door inspecteurs en toezichthouders.

Deze opdracht

In deze opdracht dient gezocht te worden naar (effectiviteits)onderzoeken met een redelijke tot hoge mate van interne validiteit (MSMS 3+). Naast een MSMS score van 3+ is er het zogenaamde matching-criterium. Hiermee bedoelen we dat de onderzoeken die worden gepresenteerd op enige wijze matchen met de wereld van handhaving en toezicht. De zoektocht moet namelijk breed ingezet worden omdat het aantal MSMS 3+ effectevaluaties op het gebied van toezicht zeer beperkt is. Matchingcriteria zijn bijvoorbeeld:

- Het onderzochte werkveld vertoont gelijkenis met het werkveld van een inspectie of handhavende instantie;
- Onderzochte subjecten zijn vergelijkbaar met bepaalde (groepen) onder toezichtgestelden / bedrijven / personen;
- Onderzochte instrumenten (zoals bijv. boetes) zijn ook van belang voor de toezicht- en handhavingswereld.

Een belangrijke doelstelling van deze studie is om te laten zien dat valide effectonderzoek wel degelijk kan. Vervolgens moet duidelijk worden gemaakt hoe dit type onderzoek te organiseren en uit te voeren. Daarnaast dient gezocht te worden naar onderzoeken die kunnen bijdragen aan de beantwoording van de effectiviteitsvraag. Hiermee bedoelen we onderzoeken met een lagere score op de MSMS-schaal (1 en 2). Gedacht kan worden aan conjoint-achtige onderzoeken waarin de counterfactual gesimuleerd wordt. Weliswaar wordt het echte effect hiermee niet aangetoond, wel wordt inzichtelijk gemaakt in welke mate relevante kenmerken van invloed zijn op het gedrag van toezichthouders of onder toezichtgestelden. Een voorbeeld is het onderzoek Fraude in Beeld dat door onderzoekers van de Bijzondere Opsporingsdiensten is uitgevoerd in 2007. Maar ook het onderzoek van Lomwel en Nelissen naar omgevingseffecten op het handhavinggedrag van gemeenteambtenaren is een voorbeeld van nuttig onderzoek met een lagere MSMS-score.

Onderzoeksvragen en
Toelichting

Algemeen gesproken dienen de onderstaande onderzoeksvragen te worden beantwoord.

Zoektocht naar effectevaluaties?

1. Welke intern valide (MSMS3+) effectonderzoeken gaan over handhaving en toezicht?
2. Welke intern valide (MSMS3+) effectonderzoeken hebben een raakvlak met de wereld van handhaving en toezicht?
3. Op welke wijze zijn de geselecteerde uitgevoerd, welk designs zijn gebruikt en wat zijn de uitkomsten?
4. Hoe zouden onderzoeken uit andere disciplines kunnen worden uitgevoerd in het veld van handhaving en toezicht?

Zoektocht naar andersoortig onderzoek dat bijdraagt aan de beantwoording van de effectiviteitsvraag

5. Welke andere onderzoeken (MSMS 1 en 2) hebben een raakvlak met de wereld van handhaving en toezicht en dragen bij aan kennis over effectief toezicht?
6. Op welke wijze zijn deze uitgevoerd, welk designs zijn gebruikt en wat zijn de uitkomsten?
7. Hoe zouden deze onderzoeken kunnen worden uitgevoerd door toezichthouders?

Spelen theorieën een rol in
het onderzoek?

Ja, zie bovenstaande tekst.

Zo ja: welke?	
Methoden en technieken (voorzien gebruik van statistische tools)	Literatuurstudie, researchsynthese waarbij ook naar internationale publicaties wordt gezocht.
Gevraagde producten en/of diensten (inclusief wijze van kennisoverdracht)	Rapport
Benodigde expertise	Toonaangevende expertise op het hier bedoelde terrein
Wijze van uitvoering	Extern
Risico's, onzekerheden, Beslispunten	-
Begeleidingscommissie	Voor dit onderzoek formeert het WODC een leescommissie bestaande uit vertegenwoordigers van WODC, ERR, Iraad en een externe expert op het gebied van toezicht.
Overig (o.m. relevante eerdere onderzoeken)	<p>Systematic research reviews</p> <p>A systematic review is a literature review usually focused on a single question which tries to identify, appraise, select and synthesize all high quality research evidence relevant to that question. It:</p> <ul style="list-style-type: none"> • addresses a clearly formulated question • uses systematic and explicit methods to identify, select and critically appraise relevant research • uses systematic and explicit methods to collect and analyze data from the studies that are included in the review • Statistical methods (meta-analysis) may or may not be used to analyze and summarize the results of included studies which are considered similar enough to combine.

Organizations like the Campbell Collaboration and the Cochrane Society produce systematic reviews. The Campbell Collaboration produces them of the effects of social interventions, ranging from crime and justice to social welfare. In a systematic review, relevant prior evaluation studies are collected and then subjected to a critical evaluation in order to provide an answer to the question which programs work (best). Petrosino et al. (2001) describe the criteria the Campbell Collaboration applies to systematic reviews. Such studies must contain detailed information on the research process, including the research question(s), the methods used when searching for and screening studies, and the criteria to be met for studies to be included. A systematic review also details how analyses were done and how conclusions were reached. Emphasis is placed on the importance of assessing the strength of research evidence when reviewing knowledge about the effects of social and behavioral interventions and programs. The point is stressed that not all evaluation designs are created in the same manner as a result of which some designs do not allow for inferring cause and effect while others do.

Whether a meta-analysis is part of a systematic review or not, systematic reviews are reviews in which rigorous methods are employed to summarize, analyze, and combine study findings (Petrosino et al., 2001). Petrosino et al. (2001: 20) note, "The foremost advantage of systematic reviews is that when done well and with full integrity, they provide the most reliable and comprehensive statement about what works." However, Petticrew (2003) discusses the limitations of this approach as and when the practical relevance of systematic reviews is taken into account. He makes the point that due to Rossi's (1987) stainless steel law of evaluation and his iron law of evaluation (which states that the expected value of any impact assessment of any large scale social program is zero) policy-makers using systematic reviews in their decision making processes, are not or only in a restricted way confronted with insights from narrative reviews and observational studies. This might jeopardize utilization and subsequently the policy-making process.

Knowledge repositories in general

Hansen & Rieper (in press) have inventoried a number of what they call (second-order) evidence-producing Organizations within the social (and behavioral)

sciences. In recent years the production of systematic reviews has been institutionalized in these institutions. There are two main second-order knowledge-producing international organizations: the Cochrane Society working within the health field; and the Campbell Collaboration working within the fields of social welfare, education and criminology. Both collaborations subscribe to the idea of producing globally valid knowledge about the effects of interventions, if possible through synthesizing the results of primary studies designed as RCTs and using meta-analysis as the form of syntheses. In many (western) countries second-order knowledge-producing organizations have been established at the national level that not all are based on findings from RCT's. Hansen & Rieper (in press) present information about some 15 of them, including web-addresses.

Bijlage t.b.v. invoer van projectgegevens in Navision en Onderzoekdatabase

Projectnummer	1761-2
Aanvrager DG	
Aanvrager directie	ERR
Contactpersoon aanvrager	Ralph Vossen
Uitvoerende afdeling WODC	EWB
Contactpersoon WODC	Willemsen
Nummer VBTB-thema	6
Opgenomen in programmering?	
zo ja: programmeringsjaar en -ronde	Ja
Begroting kosten extern uitvoeren, incl. BTW (per jaar specificeren).	2009-1
Begroting uren WODC, waarbij ook uren van EWB voor projectbegeleiding (beide per jaar specificeren).	1761-2: 35.000 Euro in 2009
Geplande begindatum project	01-03-2009
Geplande einddatum project	01-06-2009
Publicatievorm (traject aangeven)	Rapport
Aanbieding publicatie aan:	
Nummer financieringsbron ¹	1
(Mede)financiering door derden? Zo ja:	ja
1. toezegging bijvoegen	
2. door wie?	ERR
3. bedrag(en) cofinanciering	€ 35.000,-
Evaluatie	nee
Zo ja, welke?	
Kamertoezegging:	nee
zo ja: vindplaats en nummer	
Gebruik vertrouwelijke gegevens	nee
Levering databestand aan SIBa?	nee
Gebruik bron(nen) WODC?	nee
zo ja: welke?	
inzet bronbeheerder(s) (namen)	
begroting benodigde uren	

¹ Beleidsartikel (VBTB-thema)

Beleidsartikel (VBTB-thema)			Financieringsbron		
11.	(Nationale) wetgeving	14.1	Uitvoering jeugdbescherming	1	Geprogrammeerde
11.	Wetgeving kwaliteitsbeleid	14.2	Tenuitvoerlegging iustitiële	2	Handhaving
12.	Slagvaardige en kwalitatief goede rechtspleging	14.3	VERVALLEN	3	Gesubsidieerde projecten
12.	Adequate toegang tot het	15.1	Verblijfsrechten vreemdelingen	4	XTC
13.	Preventieve maatregelen	15.2	Opvang gedurende beoordeling	5	Extra projecten
13.	VERVALLEN	15.3	Terugkeer vreemdelingen		
13.	Handhaving en vervolging	16	VERVALLEN		
13.	Tenuitvoerlegging	17.1	Internationale regelgeving		
13.	Slachtofferzorg	91.1	Algemene Justitie-uitgaven		
13.	Terrorismebestrijding	91.1.	Effectieve besturing Justitie		