November 2012

Opleiding Bestuurskunde
Faculteit der Managentwetenschappen
Radboud Universiteit Nijmegen
Begeleiding: mw. dr. E. Mastenbroek

Erwin Grommen
0808164

INTERNE OF EXTERNE EVALUATIE DOOR DE EUROPESE COMMISSIE

In dit artikel wordt een onderzoek gepresenteerd naar de keuze tussen interne en externe evaluatie bij wetgevingsevaluaties in opdracht van de Europese Commissie. Zelf benadert de Commissie haar evaluatiepraktijk op een rationele, managementachtige wijze met weinig aandacht voor het politieke aspect dat aan evaluaties verbonden is. Het is echter maar de vraag of deze benadering terecht is, of dat politieke factoren nog een belangrijke rol spelen. Om dit te onderzoeken zijn hypothesen geformuleerd over de factoren die de keuze tussen een interne dan wel externe evaluatie kunnen beïnvloeden. Deze variabelen zijn de besteding van publieke middelen, het type evaluatie, de politieke gevoeligheid van de betreffende wetgeving en het werken onder tijdsdruk. Door middel van een binomiale logistische regressieanalyse is de invloed van deze variabelen onderzocht. Deze analyse liet zien dat deze variabelen van invloed zijn op deze keuze, echter niet steeds op de wijze die werd verwacht. Wat wel overeenkomt met de verwachting is dat een grotere politieke gevoeligheid de kans op een externe evaluatie vergroot. Ook de verwachting dat bij tijdsdruk de kans op een interne evaluatie groter is wordt bevestigd. Hier staat tegenover dat, in tegenstelling tot de verwachting, de kans op een interne evaluatie groter is indien een oordeel moet worden geveld, of indien de evaluatie wetgeving betreft waarbij publieke middelen zijn toegepast. Hierin verschilt de evaluatiepraktijk van de Europese Commissie dus van de gangbare theorieën. Een kritische benadering van de evaluatiepraktijk van de Commissie met aandacht voor politieke factoren lijkt dan ook gevraagd.

Inleiding

Eind jaren ’90 verkeerde de Europese Commissie in een crisis (Wille 2007, pp. 8-9). Deze crisis kwam voort uit een afnemende acceptatie en een toenemende kritiek op de legitimiteit van de Commissie (Tsakatika, 2005, pp.194, 203-204). De Europese Commissie was oorspronkelijk ontworpen als een technocratisch orgaan dat oplossingen levert voor beleidsproblemen, deals sluit, als een motor voor verdere integratie functioneert en het algemeen belang in het oog houdt. Toen er echter steeds meer vraagtekens werden gezet bij de behoefte aan verdere integratie kwam de functie van de Commissie als motor hiervoor onder druk te staan. Daarnaast was er kritiek die zich richtte op het feit dat de Commissie niet democratisch gekozen is, terwijl zij wel een zeer belangrijke politieke rol speelt (Wille 2007, pp. 8-9; Chryssochoou, 2009, pp. 379-381).

De druk op de Commissie bereikte een hoogtepunt met de val van de Commissie Santer in 1999. Deze val werd veroorzaakt door beschuldigingen van fraude, mismanagement en ongeregeldheden (van Gerven, 2000, p.6; Macmullen, 1999, pp.704-705; Tsakatika, 2005, p.203; Wille, 2007, pp. 8-9;). De positie van de Commissie ten opzichte van de lidstaten was hierdoor ernstig verzwakt. Dit is mogelijk doordat geen sprake is van vaststaande posities en bevoegdheden (Curtin & Egeberg, 2008, pp.7-8; Hoffman, 2009, p.36; Toulemonde, 1996, p.50; Rosamond, 2010, pp.115-116). In plaats daarvan worden de onderlinge verhoudingen bepaald via een politiek machtsspel (Summa & Toulemonde, 2002, p.408; Toulemonde, 1996, p.50).

Belangrijk voor de verhouding tussen de Europese Commissie als vertegenwoordiger van de EU en de lidstaten is het *subsidiariteitbeginsel.* Dit wordt omschreven als ‘*het beginsel dat er voor dient te zorgen dat beslissingen zo dicht mogelijk bij de burger worden genomen*’ (Cini & Pérez-Solórzano Borragán, 2009, p.452). Toulemonde (1996, p.49) wijst er op dat dit niet alleen betekent dat een hoger niveau niet mag handelen als een lager niveau het beter kan,

maar ook dat het hogere niveau verplicht is om te handelen als het lagere niveau het niet alleen af kan. Dit betekent dat de Commissie een toegevoegde waarde moet zijn ten opzichte van de lidstaten. Het vertrouwen in de Commissie, als boegbeeld van het hoger niveau, was echter afgenomen door het schandaal rond de Commissie Santer en haar legitimiteit kwam mede hierdoor onder druk te staan (Tsakatika, 2005, p.203). Om de positie van de Commissie weer te versterken, werden in het 'White Paper on Governance' van de Commissie Prodi pogingen gedaan om een antwoord te formuleren op de vragen over de legitimiteit van de Commissie (Tsakatika, 2005, p.195). In dit 'White Paper' worden de begrippen *efficiëntie, verantwoording, transparantie, verantwoordelijkheid* en *service* aangemerkt als kernwaarden van de Commissie (Wille, 2007, p. 11).

Via evaluaties is het mogelijk om deze waarden in de praktijk te brengen. Zij bieden immers een mogelijkheid om op transparante wijze aan burgers verantwoording af te leggen over de efficiëntie en effectiviteit van maatregelen (Barosso, 2009, p.29; Europese Commissie, 2007, p.3). Via evaluaties die een succesvol beeld van haar activiteiten geven, kan de Commissie haar toegevoegde waarde laten zien (Stame, 2008, p.126). Ook kunnen minder positieve evaluaties over de rol van lidstaten laten zien dat een grotere rol voor de Europese instituties nodig is. Evaluaties kunnen daarnaast gebruikt worden om de legitimiteit te vergroten (Derlien & Rist, 2002, p.441). Evaluaties kunnen zo, door oordelen over effectiviteit te geven, een rol spelen in verschuivingen in (opvattingen over) de taken van de Commissie en andere Europese instituties ten opzichte van de lidstaten. Ook bieden evaluaties de mogelijkheid om bestaand beleid tegen het licht te houden en eventueel van de ingeslagen weg af te wijken (Bressers & Hoogerwerf, 1991, pp. 19-21). De politieke belangen verbonden aan evaluaties zijn dus groot. Politiek verantwoordelijken kunnen dan ook in de verleiding komen om te proberen invloed uit te oefenen op het oordeel (Bovens, 't Hart & Kuipers, 2008, pp 3-4). Zelfs als de evaluatie 'an sich' neutraal en objectief is uitgevoerd, zullen de

uitkomsten politiek gebruikt gaan worden (Deetman, 1988, p.29; Wildavsky, 1972, p. 516).

De Commissie gaat in haar verhandelingen over haar evaluatiepraktijk echter niet in op de politieke mogelijkheden die evaluaties bieden (Barosso, 2009; Europese Commissie 2004, 2007, 2010). In plaats daarvan gebruikt zij een rationele, managementachtige benadering waarbij enkel op de mogelijkheden die evaluaties bieden voor het verbeteren van wetgeving en het vergroten van kennis wordt ingegaan. Of deze benadering van de evaluatiepraktijk door de Commissie terecht is, is echter maar de vraag. De politieke functie is een belangrijke reden voor de publieke aandacht voor evaluatieonderzoek. De publieke aandacht richt zich echter vooral op het evaluatieproces zelf en de resultaten achteraf.

De ‘eerste fase’ van het evaluatieproces, waarbij wordt bepaald wie de uitvoerder is, krijgt daarentegen relatief weinig aandacht. Organisaties kunnen ervoor kiezen om de evaluatie zelf uit te voeren, maar ook om de evaluatie uit te besteden. De eerste optie wordt aangeduid met de term *intern evaluatieonderzoek*, het tweede wordt *extern evaluatieonderzoek* genoemd (Herweijer & Kottman, 1988, pp. 92-93; Vedung, 1997, p.114; Worthen, Sanders & Fitzpatrick 1997, pp. 18-19). In verschillende standaardwerken over beleidsevaluatie is geen aandacht voor de vraag wie de evaluatie uitvoert (Madaus, Stufflebeam & Scriven, 1983; Mayne, Bemelmans-Videc, Hudson & Conner, 1992; Owen, 2007; Rossi, Freeman & Lipsey, 1999; Ringeling & Sorber, 1988; Swanborn, 2007; Weiss, 1972). In plaats van in te gaan op deze vraag, wordt in deze standaardwerken uitgegaan van een assumptie over de positie van de uitvoerder van evaluatieonderzoek. Vaak wordt hierbij uitgegaan van een onderzoeker van buiten de organisatie. Hiermee wordt voorbijgegaan aan het gegeven dat een opdrachtgever wel degelijk een keuze heeft. Over hoe deze keuze gemaakt wordt, is hierdoor nog weinig bekend. Een onderzoek naar deze keuze is, zeker in het geval van de Europese Commissie, wel degelijk van belang. Door onderzoek te doen naar de factoren die van belang zijn om een evaluatie al dan niet intern uit te voeren, kan worden

achterhaald of er nog een rol is voor politieke factoren naast de door de Commissie beschreven technocratische benadering van evaluatie.

In dit artikel wordt gezocht naar een antwoord op de vraag op welke factoren de keuze voor een bepaald type relatie met de onderzoeker berust bij het uitvoeren van evaluatieonderzoek in opdracht van de Europese Commissie. Deze vraag zal worden beantwoord voor de ex-post wetgevingsevaluatie. Voor dit type evaluatie is gekozen vanwege verschillende redenen. Ten eerste is wetgeving het belangrijkste beleidsinstrument van de Europese Commissie (Lodge, 2008, p.280). Ten tweede komt op het wetgevingsterrein de gecompliceerde machtsverdeling binnen de Europese Unie sterk naar voren (Hoffman, 2009, pp. 32-33). Ten slotte is de aandacht voor het ex-post evaluatieonderzoek van de Unie nog zeer beperkt, zeker vergeleken met de aandacht die er is voor andere evaluatietypen als het Impact Assessment[1].

Om de onderzoeksvraag te kunnen beantwoorden, zal de beperkte literatuur die tot nu toe over de keuze tussen een interne of externe evaluatie is verschenen nader bestudeerd worden. Hieruit worden verschillende variabelen afgeleid waarop de keuze voor een bepaalde relatie met de onderzoeker wordt gebaseerd (Conley-Tylor, 2005; Cummings e.a. 1988; Herweyer en Kottman, 1988; Vedung, 1997; Worthen e.a., 1997). De invloed van de verschillende variabelen zal via een logistische regressieanalyse worden onderzocht.

Dit artikel zal eerst ingaan op de positie van de Europese Commissie. Hierbij zal ook kort worden ingegaan op de rol van evaluaties binnen de Commissie. Vervolgens zal de gevonden literatuur worden besproken in het theoretisch kader. Hierbij wordt eerst ingegaan op het onderscheid tussen intern en extern evaluatieonderzoek. Na het achterhalen van de gebruikte argumenten zullen deze in het methodologisch kader worden geoperationaliseerd. In de hierop volgende paragraaf wordt de gebruikte dataset gepresenteerd. Ook zal de gebruikte

[1] Zie bijvoorbeeld Formez, 2004; IIEP, 2004; Kirkpatrick & Lee, 2004; Meuwese, 2008; Opoku & Jordan, 2004; Radaelli, 2005, 2007; Radaelli & Meuwese, 2010; Renda, 2006.

methode worden uitgelegd. Hierna zal de uitgevoerde analyse worden weergegeven. De afsluitende paragraaf presenteert de conclusies van het onderzoek.

De positie van de Europese Commissie en de noodzaak tot evaluatie

De omgeving waarin de Europese Commissie zich beweegt is complex. Door de verschuiving van autoriteit naar de Europese Unie is politieke fragmentatie opgetreden (Kelemen, 2006, p.105). Hierdoor vindt wetgeving in de Unie plaats in een beleidsruimte waarin via gedeelde soevereiniteit macht wordt uitgeoefend (Hoffman, 2009, pp.32-33). Enerzijds wordt het politiek gezag op een horizontaal vlak verdeeld tussen de Europese Commissie, het Europees Parlement, de Europese Raad en het Europees Hof van Justitie, anderzijds is op een verticaal vlak sprake van een verdeling van politiek gezag tussen de EU en de nationale en regionale overheden (Kelemen, 2006, p.105). Op het gebied van wetgeving komt deze gedeelde soevereiniteit bij uitstek tot uiting in de noodzaak tot transpositie. Met deze term wordt het proces aangegeven waarmee de door de EU uitgegeven wetgeving wordt omgezet in de nationale wetgeving. Hierbij moet binnen de door de EU-wetgeving vastgelegde grenzen worden gebleven (Duina, 1997, p. 156). De noodzaak tot transpositie maakt de inwerkingtreding van wetgeving een belangrijk en complex proces binnen de Europese Unie (Thatcher & Coen, 2009, 168-170). Ook in de implementatie, naleving en handhaving van de uitgegeven wetgeving spelen de nationale overheden een grote rol. De coördinatie van wetgeving is dan ook geen geringe taak. Het is de Commissie die uiteindelijk als taak heeft gekregen om een goede implementatie van wetgeving te verzekeren en zo uiteindelijk verdere integratie te bewerkstelligen (Hoffman, 2009, p.136; Thatcher & Coen, 2009, pp.170-171). Om deze taak te kunnen vervullen heeft de Europese Commissie een aantal bevoegdheden. Zo bepaalt zij in grote mate welke initiatieven van andere actoren worden omgezet in een wetsvoorstel. Ook is zij verantwoordelijk voor de monitoring van de implementatie in de lidstaten en zorgt zij voor een gedetailleerdere invulling van wetgeving

vanuit het Europees Parlement en/of de Europese Raad (Egeberg, 2009, p.127). Daarnaast is zij betrokken bij de formulering en ontwikkeling van beleid. Deze taken zijn de Commissie toebedeeld vanuit haar verantwoordelijkheid voor het nemen van maatregelen om de ontwikkeling van de EU vooruit te helpen (Nugent, 2003, p.131). Maar terwijl de EU zich ontwikkelt van een forum voor onderhandeling tussen staten naar een volwaardig politiek systeem, nemen de zorgen over het democratisch gehalte van de EU toe (Chryssochoou, 2009, pp. 379-381). Dat ook de Commissie niet aan deze kritiek ontkomt werd in de inleiding al besproken . Deze kritiek gaat gepaard met een groeiende roep om verantwoording door de Commissie aan de burgers van de EU (Chryssochoou, 2009, pp. 379-381).

Verantwoording vanuit de Commissie aan de burgers van de EU is echter eenvoudiger gezegd dan gedaan. Dit heeft alles te maken met de wisselende verhoudingen tussen de verschillende actoren en het gegeven dat de actoren van elkaar afhankelijk zijn. De democratische verantwoording in deze setting is verzwakt (Papadopuoulos, 2010, p. 1040). Dit wordt in de eerste plaats veroorzaakt door de gedeelde soevereiniteit. Ook kent de verbinding van de bureaucratie met de kiezer zoveel tussenschakels en ambtelijke vrijheid, dat de democratische verantwoording via de politiek verantwoordelijken beperkt is. Daarnaast is het door de netwerkvorm vaak moeilijk om te achterhalen wie precies waarvoor verantwoordelijk is (Papadopoulos, 2010, pp.1033-1039). Doordat de weerstand tegen een verdere politieke integratie groot is, is het bijna onmogelijk om op traditionele wijze verantwoording af te leggen aan de burger (Majone, 1998, p.21). Omdat op dit moment het draagvlak ontbreekt om hervormingen (bijvoorbeeld directe verkiezingen) door te voeren die voor een directe verbinding tussen burger en Commissie kunnen zorgen, is er grote behoefte aan andere vormen van verantwoording (pp. Curtin & Egeberg, 2009, pp.23-24). Een manier om verantwoording af te leggen en de redenen achter een beslissing inzichtelijk te maken, is het beter uitleggen van een beslissing door politici aan de burger (Majone, 1998, p.21). De

Europese Commissie ziet in evaluaties een middel om deze verantwoording aan de burger mogelijk te maken (Barosso, 2009, p.29). Ook kunnen evaluaties gebruikt worden als instrument om de implementatie te monitoren. Via evaluaties kan duidelijk worden of middelen effectief en efficiënt worden ingezet, en of dit ook leidt tot de gewenste resultaten (Rossi e.a., 1999, pp 40-42; Vedung, 1997, pp 102-103).

Dit neemt echter niet weg dat evaluaties ook misbruikt kunnen worden. Het komt voor dat een evaluatie enkel wordt gestart omdat de opdrachtgever denkt dat een evaluatie voor goede publiciteit zal zorgen of indruk zal maken op financiers. Ook kan het zijn dat een evaluatie wordt uitgevoerd om een context te creëren voor een te nemen beslissing die al vaststaat. Daarnaast kan een evaluatie dienen als vertragingstactiek. Dergelijke redenen spelen in praktisch alle evaluaties in meer of mindere mate een rol (Rossi e.a., 1999, pp 42-43; Vedung, 1997, pp 111-113). De resultaten die een opdrachtgever met een evaluatie wil bereiken kunnen dus zeer divers zijn.

Interne of externe evaluatie

Om de gewenste resultaten te bereiken zal de opdrachtgever de juiste keuze moeten maken voor een uitvoerder. Dit betekent natuurlijk dat voor de juiste personen gekozen moet worden, maar voordat deze keuze gemaakt kan worden zal een fundamentelere keuze gemaakt moeten worden: voert de organisatie de evaluatie zelf uit, of wordt een onderzoeker van een andere organisatie ingehuurd?

Naar deze keuze wordt gerefereerd met de termen intern en extern. Met deze termen wordt een verschil aangegeven tussen de verschillende organisatorische posities die een evaluatieonderzoeker kan innemen ten opzichte van de opdrachtgevende instantie (Herweijer & Kottman, 1988, pp. 92-93; Vedung, 1997, p.114; Worthen e.a., 1997, pp. 18-19).

Volgens een algemeen geaccepteerd onderscheid wordt met de term extern evaluatie-onderzoeker bedoeld dat de onderzoeker niet in dienst is van de organisatie die

verantwoordelijk is voor het beleid (Herweijer & Kottman, 1988, pp. 92-93; Vedung, 1997, p.114; Worthen e.a., 1997, pp. 18-19). Met de term intern evaluatieonderzoeker wordt juist wel bedoeld dat de onderzoeker in dienst is van de organisatie verantwoordelijk voor het te evalueren beleid. In de praktijk is dit onderscheid echter minder zwart-wit weer te geven (Herweijer & Kottman, 1988, p. 92-93; Mathison, 1994, p.300). Er zijn verschillende mogelijkheden wat betreft de organisatorische positie van met name interne evaluatieonderzoekers. Deze kunnen verspreid zijn over verschillende beleidsdirecties, maar ook samengebracht zijn in een min of meer onafhankelijke afdeling. Op dit onderscheid zal hier echter niet worden ingegaan, intern blijft immers intern, hoe dit ook is georganiseerd. Enkel het algemeen erkende onderscheid tussen intern en extern onderzoek zal centraal staan.

Dit onderscheid staat pas vrij recent in de belangstelling. Lange tijd was het namelijk vanzelfsprekend dat evaluatieonderzoek extern plaatsvond. Pas vanaf ongeveer 1980 was hierin een kentering zichtbaar en groeide ook de populariteit van het intern evaluatieonderzoek (House, 1986, pp. 63-64; Love, 1983, p.1). Dit kan een verklaring zijn voor het feit dat het grootste deel van de beschikbare standaardwerken voorbij gaan aan de keuze tussen het intern of extern uitvoeren van evaluatieonderzoek, maar er zonder echte verantwoording van uitgaan dat het onderzoek extern uitgevoerd wordt. Ondanks het feit dat er slechts weinig aandacht voor deze keuze is, is het door het vergelijken van diverse auteurs mogelijk een aantal hypothesen te formuleren wanneer een intern dan wel een extern evaluatieonderzoeker kan worden verwacht[2].

Het doel van het onderzoek

Er is slechts één variabele die in alle literatuur terugkomt. Niet alleen Conley-Tylor (2005, pp. 8-9) maar ook Herweijer & Kottman (1988, pp. 96-98), Vedung (1997, pp. 116-

[2] Hoewel deze literatuur is geschreven voor beleidsevaluatie zal zij in dit artikel worden toegepast op wetgevingsevaluatie. Hiervoor is gekozen vanwege de zeer beperkte beschikbare literatuur, en het gegeven dat wetgeving het voornaamste beleidsinstrument is dat de Europese Commissie ter beschikking staat (Lodge, 2008, p.280).

119) , Worthen e.a. (1997, pp. 14, 18-20) en Cummings e.a. (1988, pp.61-62,72) geven aan dat het doel van het onderzoek van invloed is op de keuze voor interne of externe evaluatie. Zij gebruiken deze variabele echter niet steeds op dezelfde wijze. Er is sprake van kleine verschillen. Deze verschillen worden duidelijk tijdens een stapsgewijze behandeling van de verschillende beschrijvingen van deze variabele door de auteurs.

Conley-Tylor (2005, pp. 8-9) geeft aan dat intern evaluatieonderzoek de gebruikelijke vorm is voor onderzoek dat moet leiden tot veranderingen in de organisatie en het stroomlijnen en verbeteren van interne processen. Dit omdat zij over meer kennis over de organisatie en het (wetgeving)programma beschikken.

Een extern evaluatieonderzoek wordt gebruikt (Conley-Tylor, 2005, p.8) indien het onderzoek het verantwoorden van het gebruik van publieke middelen tot doel heeft. Dit omdat een extern evaluatieonderzoeker door een grotere vermeende neutraliteit voor een grotere legitimiteit zorgt.

Herweijer en Kottman (1988, p. 97) zien net als Conley-Tylor dat een interne evaluatie het meest gebruikt wordt bij onderzoek dat is gericht op het werkproces van de beleidsuitvoerder. Het doel van dit type onderzoek is om informatie te verkrijgen die kan worden gebruikt om het programma te verbeteren. Omdat een interne evaluatieonderzoeker over meer voorkennis van de wetgeving en haar context beschikt, is deze geschikter voor een dergelijke evaluatie.

Als het evaluatieonderzoek echter als doel heeft, om een oordeel te geven over de effectiviteit van de wetgeving, kan een extern evaluatieonderzoeker worden verwacht. Dit omdat de informatie die voor dit onderzoek nodig is, vaak over processen gaat die niet direct te beïnvloeden zijn door het ambtelijk apparaat. In dergelijke gevallen is vaak sprake van een gebrek aan kennis onder het eigen personeel. Daarnaast kan een extern evaluatieonderzoeker als een neutrale buffer dienen (Herweijer & Kottman, 1988, p 97-98).

Ook Vedung (1997, pp. 118-119) onderschrijft de keuze voor een intern evaluatie-onderzoeker op basis van de wens om te komen tot een verbetering van interne processen. Hiervoor gebruikt hij soortgelijke argumenten. Wel legt hij meer nadruk op het snelheidsvoordeel dat de interne evaluatieonderzoeker heeft. Door diens voorkennis van het beleid en de organisatie kan het onderzoek sneller worden gestart.

Vedung (1997, pp. 117-118) ziet de externe evaluatie vooral plaatsvinden in een setting waarbij via de evaluatie verantwoording wordt afgelegd aan derden. Verder laten ook Cummings e.a. (1988, pp. 61-62), in een onderzoek vanuit het bedrijfsleven, zien dat interne evaluatieonderzoekers zeer geschikt worden gevonden voor het uitvoeren van onderzoek gericht op de verbetering van werkprocessen door het doen van aanbevelingen.

Als de evaluatie tot doel heeft de effectiviteit van (wetgeving)programma's te beoordelen, dan geven ook Cummings e.a. (1988, pp. 61-62, 72) aan dat de praktijk leert dat dan beter met extern evaluatieonderzoekers gewerkt kan worden. Dit omdat in dit geval loyaliteitsproblemen kunnen optreden bij intern evaluatieonderzoekers die hun collega's niet willen afvallen. Daarnaast is het vertrouwen van potentiële derden in een positieve intern uitgevoerde evaluatie minder dan het vertrouwen van derden in een positieve extern uitgevoerde evaluatie.

Tenslotte signaleren ook Worthen e.a. (1997, pp. 14, 19-20) dat het type onderzoek een belangrijke indicator is voor het meest geëigende type evaluatieonderzoeker. Intern evaluatieonderzoek is zeer geschikt voor het doen van formatief evaluatieonderzoek, onderzoek gericht op het verbeteren van een (wetgeving)programma. Dit vanwege de grotere kennis die een intern evaluatieonderzoeker over het (wetgeving)programma heeft. Een mogelijk gebrek aan objectiviteit is hierbij niet zo'n probleem. Dit is anders in het zogenoemde summatief evaluatieonderzoek. Dit type onderzoek wordt uitgevoerd om een oordeel te vellen over de waarde van een programma in relatie tot bepaalde criteria . Dit

onderzoek kan vanwege het grote belang dat hierbij aan objectiviteit van het onderzoek wordt gehecht, beter extern worden uitgevoerd.

Dit overzicht kan kort worden samengevat door te stellen dat het intern onderzoek de meest voorkomende vorm is indien het doel van het onderzoek is, om te komen tot het verbeteren van wetgeving. De externe evaluatie wordt vooral gebruikt als het doel van de evaluatie het beoordelen van een beleid is.

> Hypothese I: Indien het doel van de evaluatie is om te komen tot een beoordeling van het succes of falen van het (wetgeving)programma, is de kans op een externe evaluatie groter dan indien sprake is van een evaluatie met als doel het bijdragen aan de verbetering van de processen die nodig zijn om tot een volledige invoering van de wetgeving te komen.

Besteding van publieke, financiële, middelen

Conley-Tylor (2005, p.8) stelt dat evaluatieonderzoek naar de besteding van publieke middelen altijd het beste kan worden uitgevoerd door externe onderzoekers. Omdat objectiviteit hierbij van groot belang is, dient een andere dan de verantwoordelijke organisatie de evaluatie uit te voeren (Vedung, 1997, p.118). Zeker in het geval waarbij publieke middelen worden uitgegeven is dit van groot belang. De evaluatie vindt dan namelijk plaats in het publieke domein (Conley-Tylor, 2005, p.8). Er wordt dan gekozen voor de variant waarbij de legitimiteit van de evaluatie naar derden het hoogst is, dit is de externe evaluatie.

> Hypothese II: Indien de evaluatie betrekking heeft op wetgeving waarbij de besteding van publieke middelen in de wetgeving expliciet aan de orde is gekomen, is de kans op een externe evaluatie groter, dan wanneer er niet expliciet sprake is van de besteding van publieke gelden in de wetgeving.

Tijdsdruk

De keuze van een intern evaluatieonderzoeker is aantrekkelijk indien op korte termijn een evaluatie moet worden verricht. Indien eerst nog een extern onderzoeker moet worden geselecteerd, gaat veel kostbare tijd verloren. De intern onderzoeker is al aanwezig, beschikt over meer voorkennis en kan direct aan het werk gezet worden (Conley-Tylor, 2005, p. 6; Herweijer en Kottman, 1988, p.98; Weiss, 1972, p.35).

> Hypothese III: Indien een evaluatie onder tijdsdruk moet plaatsvinden, is de kans op een interne evaluatie groter, dan indien geen sprake is van tijdsdruk.

Politieke gevoeligheid

Opdrachtgevers kunnen in de verleiding komen om feiten en beelden te creëren die overeenkomen met de eigen doelen (Bovens, 't Hart & Kuipers 2008, pp. 3-4). Deze verleiding om invloed uit te oefenen op de uitkomsten van een evaluatie is groter bij politiek gevoelige onderwerpen. Het risico bestaat dat de evaluatie enkel wordt uitgevoerd om een al vaststaande beslissingen te legitimeren (Owen, 2007, p.141). Vaak wordt gedacht dat de mogelijkheden hiervoor bij intern evaluatieonderzoek groter zijn, dan bij extern onderzoek. Het gelijkstellen van intern onderzoek aan partijdig onderzoek en extern onderzoek aan onafhankelijk onderzoek is echter voorbarig. Hierbij spelen meer factoren een rol (Herweijer en Kottman, 1988, p.99).

Extern onderzoekers kunnen bijvoorbeeld uit angst om een groot opdrachtgever te verliezen en/of toekomstige opdrachten in gevaar te brengen, net zozeer geneigd zijn om de opdrachtgever naar de mond te praten (Conley-Tylor, 2005, p.8). Intern evaluatieonderzoekers kunnen zelfs worden gezien als beter beschermd tegen represailles bij onwelkome conclusies omdat zij een grotere wettelijke bescherming genieten. Ook Cummings e.a. (1988, p.62) menen dat een extern evaluatieonderzoeker door een grote klant net zozeer onder druk kan worden gezet om wenselijke resultaten te produceren, als een intern evaluatieonderzoeker. Als de schijn van vooringenomenheid moeten worden voorkomen, wat

zeker het geval is bij politiek gevoelige onderwerpen, doet een organisatie er echter goed aan voor een extern evaluatieonderzoeker te kiezen (Conley-Tylor, 2005, pp.8-9).

> Hypothese IV: Indien een evaluatie betrekking heeft op politiek gevoelige wetgeving, is de kans op een externe evaluatie groter, dan indien sprake is van politiek minder beladen wetgeving.

Aantal evaluaties in een periode

Een andere variabele die de kans op een keuze voor extern evaluatieonderzoek kan beïnvloeden, is het aantal evaluaties dat in een periode moet worden uitgevoerd. Deze factor wordt echter gebruikt om zowel een groter gebruik van interne evaluaties, als om een groter gebruik van externe evaluaties te verklaren.

Evaluatiecapaciteit van de Europese Commissie wordt bepaald door verschillende factoren. Zo zijn de beschikbare middelen voor evaluatie, maar ook de menskracht die de interne evaluatieafdelingen worden toebedeeld van belang. Beiden zijn niet oneindig. In een periode waarin de vraag naar evaluaties groot is, kan de interne evaluatieafdeling overvraagd worden. Om dit op te lossen, kan er voor gekozen worden om externen in te huren en zo een tijdelijke capaciteitsverhoging te realiseren (Herweijer en Kottman, 1988, p.98).

Het inhuren van externe evaluatieonderzoekers is over het algemeen echter een kostbare gelegenheid. Tegen het capaciteitsargument is dan ook in te brengen dat juist als er veel evaluaties uitgevoerd dienen te worden, deze intern zullen plaatsvinden. Per uur zijn interne evaluatieonderzoekers vaak aanzienlijk goedkoper dan externe evaluatieonderzoekers. Hiervoor is het wel noodzakelijk dat voldoende evaluaties plaatsvinden om het in stand houden van de interne evaluatiecapaciteit te kunnen verantwoorden (Conley-Tylor, 2005, p. 6; Herweijer & Kottman, 1988, p.98; Worthen e.a. 1997, pp. 19-20).

In dit geval lijkt het echter waarschijnlijker dat de grootte van de evaluatie-unit van doorslaggevend belang is. De Europese Commissie beschikt vergeleken met andere

overheden immers maar over een beperkte personele capaciteit (Toulemonde, 1996, pp.51-52). Dit leidt tot de volgende hypothese:

> Hypothese V: hoe meer evaluatieonderzoeken een Directoraat-Generaal (DG) in een jaar wil uitvoeren des te groter de kans dat door de beperkte interne capaciteit wordt gekozen voor een extern evaluatieonderzoek.

Kans op het gebruik van een evaluatie

De kans dat een evaluatie benut wordt, is groter bij intern evaluatieonderzoek. Althans, tot deze conclusie komen Herweijer en Kottman (1988, p.98). Dit zou met name veroorzaakt worden doordat de wereld van beleid en de wereld van het onderzoek ieder hun eigen taal spreken. Om te voorkomen dat de communicatiekloof te groot wordt, zou voor intern evaluatieonderzoek gekozen moeten worden.

Deze conclusie wordt afgezwakt door Conley-Tylor (2005, pp. 8). Zij stelt dat de vaardigheid om relevante en tijdige informatie te communiceren met de opdrachtgever net zozeer een voordeel is van de intern evaluatieonderzoeker, als dat het een deel van de opdracht en vaardigheden is voor de extern evaluatieonderzoeker. Zij (2005, p.9) geeft ook aan dat interne evaluatieonderzoekers vooral een voordeel genieten bij het vergroten van de kans op gebruik van evaluaties bedoeld voor interne verbeteringen. Bij evaluaties bedoeld voor verspreiding buiten de organisatie is het voordeel van interne onderzoekers minimaal of zelfs afwezig. Een mogelijke verklaring hiervoor is dat in een interne evaluatie, juist doordat zij dezelfde ‘taal’ spreken, meer jargon gebruikt wordt.

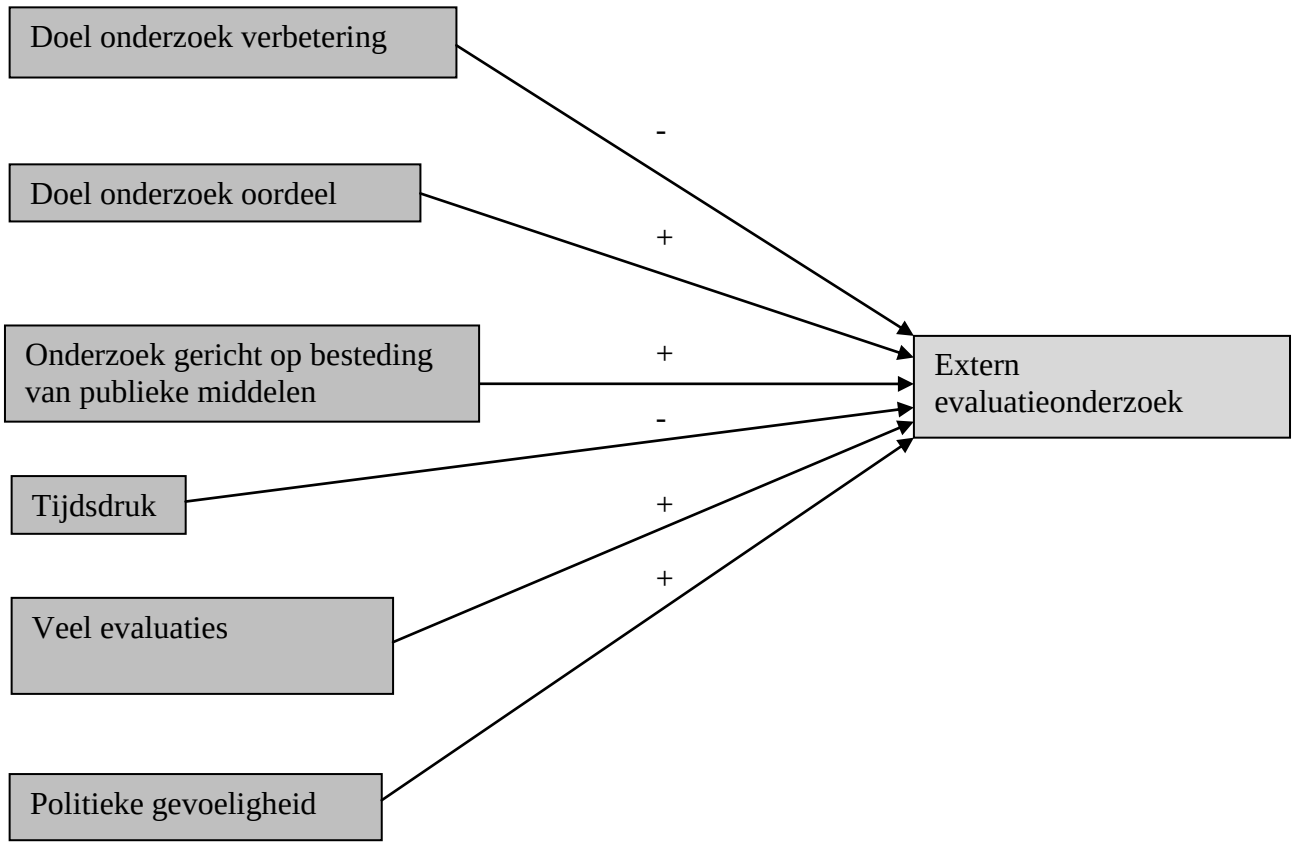
Cummings e.a. (1988, p.72) zien dat het gebruik van een evaluatie alleen in gevaar komt als er sprake is van conflicterende rollen. Bijvoorbeeld als een intern evaluatieonderzoeker het programma van een collega moet beoordelen. Er wordt gesteld dat bij een interne evaluatie slecht nieuws minder snel via de officiële kanalen naar de top gaat omdat er angst voor represailles kan zijn.

Dit in ogenschouw nemende, lijkt het er op dat bruikbaarheid geen reden is om voor intern of extern onderzoek te kiezen. De bruikbaarheid van intern onderzoek lijkt groter als het gaat om de verbetering van interne processen terwijl de bruikbaarheid van extern onderzoek groter is, als er een oordeel over een programma moet worden gegeven. Het 'bruikbaarheidargument' ligt dan in het verlengde van het 'type onderzoek argument'.

De conclusie dat de bruikbaarheid van de evaluatie geen argument is om te kiezen voor intern of extern wordt versterkt door een korte blik op de methoden die Vedung (1997, pp. 280-287) onderscheidt om het gebruik van evaluatieonderzoek te vergroten. Diens oplossingen verwijzen naar de presentatie van de bevindingen in het evaluatierapport, het besteden van aandacht aan de communicatie van de resultaten, eigenschappen van de ontvangende organisatie en het bundelen van resultaten uit verschillende evaluaties om zo meer inzicht te krijgen in wat wel en niet werkt. De wijze waarop de onderzoeker verbonden is aan de opdrachtgever wordt hierbij niet als een factor van belang beschouwd.
Dit alles in overweging nemende wordt deze factor niet beoordeeld als een factor die daadwerkelijk van invloed is. Er wordt dan ook geen hypothese opgesteld.

Aan de hand van dit overzicht kan een causaal model (*Figuur 1* op de volgende pagina) worden opgesteld waarin de verschillende factoren waarvan wordt verwacht dat ze invloed hebben, zijn opgenomen. De richting van het verband is achter het argument vermeld. Steeds is hierbij uitgegaan van de invloed op de kans op een externe evaluatie. Dit om het overzicht enigszins te behouden. Immers, wordt de kans op een extern onderzoek kleiner geacht, betekent dit automatisch dat de kans op een intern onderzoek groter wordt geacht en visa versa.

Figuur 1: causaal model



Methodologie

Het meten van de invloed van variabelen zal op een kwantitatieve manier gebeuren. Voor deze benadering is gekozen omdat in relatief korte tijd een groot aantal onderzoekseenheden moet worden bekeken, Een grote N is noodzakelijk voor de externe validiteit van het onderzoek. Er zijn echter slechts beperkte middelen beschikbaar waardoor bijvoorbeeld interviews en het meelopen op een evaluatieafdeling niet haalbaar zijn. Een analyse van verschillende evaluaties aan de hand van kwantificeerbare indicatoren is wel mogelijk.

Om de analyse uit te kunnen voeren, is een dataset samengesteld. Hiervoor zijn 169 interval en ex-post wetgevingevaluaties uit de periode 2000 tot en met 2011 geselecteerd waarvoor de Europese Commissie opdrachtgever is. Voor deze periode is gekozen vanwege de betere digitale beschikbaarheid van deze evaluaties. In de dataset wordt verder geen onderscheid gemaakt tussen ex-post en interim evaluaties aangezien wetgeving over het

algemeen geen einddatum heeft waardoor ook de ex-post evaluaties nog tijdens het bestaan van de wetgeving worden uitgevoerd.

Omdat de variabelen die van invloed zijn op de keuze voor ex-post evaluaties, worden onderzocht, kunnen enkel cases worden opgenomen waarbij ook sprake van een keuze was. Dat betekent dat voor ieder DG is nagegaan of het ook de beschikking heeft over evaluatiecapaciteit. Dit is gebeurd door de organogrammen van de DG's te bestuderen en de Annual Overviews te doorzoeken op interne evaluaties uitgevoerd door de verschillende DG's. De geselecteerde evaluaties betreffen zowel richtlijnen, verordeningen als beslissingen[3].

Voor het achterhalen van evaluaties is gebruikgemaakt van verschillende bronnen. Zo zijn de Annual Overviews[4] van de jaren 2000 tot en met 2009 doorzocht. Deze overzichten zijn door de Europese Commissie op haar website gepubliceerd. De overzichten bleken echter niet compleet. Dit bleek bijvoorbeeld toen een lijst van onderzoeksbureau Eureval[5] is doorzocht op nieuwe cases. Verdere aanvulling is ook verkregen door systematisch op zoek te gaan naar evaluaties van richtlijnen die in de periode 2000 tot en met 2004 in werking zijn getreden. Dit is gebeurd door via Google te zoeken op de betreffende richtlijnen. Ook hierbij zijn nog nieuwe cases te voorschijn gekomen. Nog steeds bleek echter na controle van de evaluatiebepalingen van de verordeningen uit 2000 dat de dataset niet compleet was. Dit betekent dat geen zekerheid kan worden gegeven over de volledigheid van de gebruikte dataset. Er zijn geen aanwijzingen voor het systematisch ontbreken van evaluaties. De

[3] Richtlijnen leggen specifieke, bindende doelen op aan lidstaten die voor een bepaalde datum moeten worden bereikt. Hoe de doelen worden bereikt, wordt overgelaten aan de lidstaat. Dit is hun discretionaire bevoegdheid. Verordeningen zijn algemeen toepasbaar en hoeven niet eerst geïmplementeerd te worden om van kracht te zijn. Zij zijn dus direct toepasbaar. Besluiten tenslotte zijn gericht op individuen en zijn in hun geheel bindend.(Kapsis, 2010, p.182).

[4] Dit zijn overzichten die worden opgesteld door het Directorate General for the Budget. Hierin is per jaar per DG aangegeven welke evaluatieactiviteiten zijn uitgevoerd. Gebleken is echter dat deze overzichten verre van compleet zijn.

[5]Dit bureau staat aan de basis van het Centre for European Expertise and Evaluation, het eerste Franse onderzoekscentrum voor de evaluatie van overheidsbeleid. Dit centrum is in Lyon gevestigd sinds 1994 (Euréval, 2012).

evaluaties zijn verspreid over de gehele tijdsperiode. Ook zijn de evaluaties verspreid over de verschillende DG's. Dit betekent niet dat bias is uit te sluiten, maar een systematische bias niet te verwachten is. Uit deze dataset zal dan ook worden getracht een model te construeren.

De eerste methode die gebruikt kan worden voor het construeren van een model waarbij een te verklaren variabele en een aantal verklarende variabelen met elkaar worden verbonden is een lineaire regressieanalyse (McClave, Benson & Sincich, 2007, 411-413). Een lineaire regressieanalsyse gaat er echter van uit dat de afhankelijke variabele continu van aard is. De positie van de onderzoeker is echter niet continu van aard, maar categorisch. Een van de onderzoeksmethoden die dan wel te gebruiken is, is de logistische regressieanalyse. Logistische regressieanalyse biedt de mogelijkheid om de invloed van een aantal onafhankelijke variabelen op een afhankelijke variabele van nominale aard met slechts een of twee categorieën na te gaan (Field, 2009, p.265; Sieben, 2002, n.p.). Binnen de logistische regressieanalyse zijn er echter verschillende mogelijkheden.

In dit artikel zal gebruik worden gemaakt van de zogenoemde Backward LR methode. Deze methode (Field, 2009, p.272) berekent eerst een basismodel waarin alle verklarende variabelen zijn opgenomen. Vervolgens wordt berekend of een van de variabelen kan worden verwijderd zonder dat de verklarende waarde van het model substantieel wordt aangetast. Hiervoor wordt een 'likelihood ratio' statistiek gebruikt. Deze methode maakt het dus mogelijk om het effect van verschillende hypothesen te meten en te vergelijken.

Omdat een statistische meetmethode is gekozen, zal de dataset op een numerieke schaal meetbaar moeten zijn. Dit betekent dat kwalitatieve verklarende variabelen zijn omgezet in dummyvariabelen om het mogelijk te maken deze op een numerieke schaal te meten (McClave, Benson & Sincich, 2007, p.448). Dit maakt het belang van de operationalisatie alleen maar groter.

De afhankelijke variabele 'type onderzoek' kent twee mogelijke uitkomsten. Een

onderzoek kan zowel door het DG zelf, intern, of door ingehuurde derden, extern, worden uitgevoerd. In de evaluatiepraktijk van de Europese Commissie bestaat echter ook de variant 'intern met externe hulp'. Deze gecombineerde vormen zijn steeds individueel beoordeeld. Hierbij is gekeken naar de rol van de externe onderzoeker. Waar deze puur ondersteunend was, is het onderzoek als intern onderzoek beoordeeld. Is het onderzoek daarentegen duidelijk en ondubbelzinnig door de extern evaluatieonderzoeker uitgevoerd met interne ondersteuning, is het onderzoek als extern aangemerkt. Waar geen duidelijkheid is over de taakverdeling, wordt aangenomen dat zowel de intern als de extern evaluatieonderzoek voor het gehele onderzoek hebben samengewerkt. Een dergelijk geval zal worden beschouwd als een intern evaluatie.

Voor het bepalen van de eerste onafhankelijke variabele, het doel van de evaluatie, kan worden geput uit de evaluaties zelf. Hiervoor is de focus van de evaluatie bepaald. Enerzijds zijn er de evaluaties die zich richten op de verschillende fasen in het proces dat moet leiden tot functionerende wetgeving, anderzijds zijn er evaluaties die zich richten op het bepalen van het succes van de wetgeving. In dit onderzoek worden met evaluaties over het proces dat moet leiden tot functionerende wetgeving evaluaties bedoeld, waar het doel is om problemen met bijvoorbeeld transpositie, invoering, uitvoering, handhaving en/of naleving van de wetgeving op te sporen. Met evaluaties voor het bepalen van het succes worden in deze context evaluaties bedoeld waarbij wordt geprobeerd een oordeel te geven over de efficiëntie, effectiviteit, doelbereiking en/of eventuele andere effecten van de wetgeving. Deze informatie is verkregen door de doelstelling van de evaluatie te achterhalen die in de inleiding of conclusie wordt vermeld.

Ten tweede is onderzocht of de evaluatie betrekking heeft op wetgeving waarbij publieke financiële middelen worden toegezegd. Hiervoor zijn de wetgevingsartikelen die geëvalueerd worden onderzocht. Als hierin een artikel is opgenomen waarin publieke

financiële middelen worden toegezegd, telt de evaluatie als een onderzoek naar wetgeving waarbij sprake is van besteding van publieke financiële middelen.

De derde onderzochte variabele is tijdsdruk. Deze variabele wordt geoperationaliseerd door de initiator van de evaluatie te achterhalen. Een groot deel van de evaluaties die de DG's in opdracht van de Europese Commissie uitvoeren, zijn het gevolg van een bepaling in een wet die al jaren vastligt. Hierbij zal geen plotselinge tijdsdruk optreden. Het kan echter ook zijn dat evaluaties worden uitgevoerd om een andere reden. Bijvoorbeeld omdat de Europese Commissie of het Europees Parlement informatie nodig hebben. Het kan zijn dat dan minder tijd beschikbaar is, om hieraan tegemoet te komen en dat tijdsdruk hier wel een rol speelt. Natuurlijk zal niet iedere evaluatie die niet vanuit een evaluatiebepaling in de wet wordt gestart plotseling en op korte termijn moeten plaatsvinden, wel zal het moeilijker zijn, hierop te anticiperen.

Ten vierde is er de variabele capaciteit. Hiervoor zou eigenlijk bekend moeten zijn wat de beschikbare mankracht is en welke middelen per jaar per DG beschikbaar zijn. Ook zou het beslag dat een evaluatie zou leggen op de mankracht en de middelen hiervoor bekend moeten zijn. Idealiter zouden deze gegevens zowel voor de interne als de externe variant bekend zijn. Ook de andere activiteiten die door de interne unit moeten worden uitgevoerd zouden bekend moeten zijn, deze zal immers niet alleen ex-post wetgevingsevaluaties uitvoeren maar ook bijvoorbeeld impact assessments en evaluaties van niet-wetgeving. Het verzamelen van deze informatie is echter niet mogelijk. De invloed van de variabele capaciteit lijkt dan ook niet te meten binnen de mogelijkheden van dit onderzoek. Dit betekent dat het ontstaan van vertekening niet kan worden uitgesloten. Het effect van deze vertekening is lastig in te schatten. Desondanks blijft het mogelijk om gerichte uitspraken te kunnen doen over de invloed van variabelen. Zolang er namelijk geen enkel DG is dat enkel interne of enkel externe evaluaties uitvoert, blijft de assumptie dat het DG een keuze heeft gemaakt voor

ofwel een externe dan wel een interne evaluatie gerechtvaardigd. Voor deze keuze is dan een afweging gemaakt gebaseerd op andere variabelen dan alleen capaciteit.

Tenslotte is er de variabele politieke gevoeligheid. Hoewel deze moeilijk te meten is, is toch een poging gedaan. Hiervoor zal het aantal amendementen van de wetgeving die onderwerp is van de evaluatie worden onderzocht. Dit is een indicatie voor het plaatshebben van een politiek debat voorafgaand aan de vaststelling. In cases waarbij meerdere wetgeving als basis wordt gebruikt, is de wetgeving met het hoogste aantal amendementen leidend. Daarnaast zijn er bepaalde sectoren van EU-wetgeving die vooraf al gevoeliger liggen dan andere sectoren. Dit komt onder andere terug in de besluitvormingsprocedure van de Europese Unie. Op gevoelige beleidsterreinen is het namelijk zo, dat ieder land een vetorecht heeft. Op minder gevoelige beleidsterreinen worden besluiten genomen via een gekwalificeerde meerderheid. Als politiek gevoelige terreinen worden het buitenlands- en veiligheidsbeleid, sociale zekerheid, belastingen, ontslagrecht, belastingen en de begroting beschouwd (Princen, 2010, pp.42 - 48). Deze terreinen vallen grotendeels onder het DG 'foreign policy instruments service', het DG 'enlargement', het DG 'employment, social affairs and inclusion', het DG 'justice and home affairs', het DG 'Budget' en het DG 'taxation and custom union' (Van Voorst, 2012, p.51). Evaluaties die horen bij deze departementen zullen in het onderzoek dan ook als politiek gevoelig worden beschouwd. Om dit mee te rekenen zal aan een dergelijke evaluaties een extra score van 10 worden toegekend. Dit om aan te geven dat het terrein gevoeliger is, maar dat het debat nog steeds als leidend wordt beschouwd. De waarde van 10 wordt als geschikt hiervoor beschouwd, omdat het gemiddeld aantal amendementen ongeveer 37 bedraagt. Een toevoeging van 10 betekent dat wetgeving met 37 amendementen op een gevoelig beleidsterrein ongeveer 25% gevoeliger wordt beschouwd als wetgeving met 37 amendementen op een niet gevoelig beleidsterrein. De invloed van het beleidsterrein wordt zo inzichtelijk, terwijl het aantal amendementen toch leidend blijft.

In dit onderzoek worden dus vier onafhankelijke variabelen onderzocht wat dit onderzoek een multivariate analyse maakt (Vennix, 2007, p.241). De te onderzoeken variabelen zijn in tabel 1 op deze pagina weergegeven.

Tabel 1: overzicht van de te onderzoeken variabelen

Theoretische variabele	Indicator	Meetniveau	Scores	Type	Verwachte kans op extern onderzoek
Positie onderzoeker	N.V.T.	Nominaal	Intern=0 Extern =1	Afhankelijk	
Type evaluatie	Evaluandum: oordeel of proces	Nominaal	Proces=0 Oordeel=1	Onafhankelijk	+
Besteding publieke gelden	In wetgeving toezegging van publieke gelden	Ordinaal	Nee=0 Ja=1	Onafhankelijk	+
Tijdsdruk	Aanleiding: wet of door EP/EC	Nominaal	Door wet =0 Door EP/EC/Raad =1	Onafhankelijk	+
Politieke gevoeligheid	Aantal amendementen & gevoeligheid beleidsterrein	Ratio	N.v.t.	Onafhankelijk	+

Analyse

Zoals tabel 2 op de volgende pagina laat zien, zijn voor dit onderzoek in totaal 169 ex-post wetgevingevaluaties bekeken. De methode ‘logistische regressie’ laat echter geen cases toe die informatie missen. Er zijn dan ook verschillende cases die door SPSS uit de analyse zijn verwijderd omdat deze op ten minste één variabele informatie missen. Met name de ‘Aanleiding’ en de ‘Politieke gevoeligheid’ waren met respectievelijk 58 en 48 missing cases voor veel evaluaties niet te achterhalen. Van de overgebleven 78 evaluaties zijn 49 (62,8%) evaluaties extern uitgevoerd, terwijl 29 (37,2%) evaluaties intern zijn uitgevoerd. Dit betekent dat de verhouding tussen intern en extern is veranderd ten opzichte van de complete database. Van de 169 gevonden evaluaties werden er 75 (44,4%) intern uitgevoerd. 94 evaluaties (55,6%) werden extern uitgevoerd.

Tabel 2: Casusoverzicht

	N	Percentage	Relatief
In analyse	78	46,2	100
Intern	*29*	*17,2*	*37,2*
Extern	*49*	*29,0*	*62,8*
Missing	91	53,8	
Niet geselecteerd	0	0	
Totaal	169	100	

Nu kan met de daadwerkelijke analyse met de methode *Backward LR* worden begonnen. Voordat de analyse kan worden geïnterpreteerd, moet een residu-analyse worden uitgevoerd om de kwaliteit van het model te kunnen beoordelen en eventueel te verbeteren. Een residuanalyse in de logistische regressie richt zich niet op het controleren van aannamen (Lammers, Pelzer, Hendrickx & Eisinga, 2007, p.161). Waar het wel om gaat, is tweeledig (Field, 2009, p.292). Het eerste doel is om casussen te identificeren waarvoor het model niet voldoet. Het tweede is om casussen te identificeren die een onevenredige invloed op het model uitoefenen. De resultaten van de residuanalyse zijn weergegeven in tabel 3 op de volgende pagina. Casussen met een gestandaardiseerde residu-waarde van boven de 2,58 zijn hierin opgenomen. Een waarde van meer dan 2,58 wijst namelijk op significante afwijkingen (Field, 2009, p.293; Lammers e.a., 2007, p.162). Hoewel het verleidelijk is dergelijke cases te verwijderen en zo de fit van het model te vergroten, is dit niet gerechtvaardigd (Field, 2009, p.293). Deze casussen zullen juist extra goed moeten worden bestudeerd om de afwijkingen van het model te verklaren. Het is echter wel mogelijk dat het noodzakelijk is casussen te verwijderen. Om dit te bepalen kan de *Cook distance* worden gebruikt. Deze geeft de invloed van het individu op de eigen fit en die van andere individuen aan (Field, 2009, p.217; Lammers e.a., 2007, p.162). Deze waarde hoort onder de 1 te liggen. Is dit niet het geval, dan is de kans op vertekening van het model erg groot (Field, 2009, p. 293). In tabellen 3a en 3b op de volgende pagina zijn de resultaten van deze analyse weergegeven.

Tabel 3a: residuanalyse

Casus-nummer	Geobserveerde waarde	Voorspelde waarde	Voorspelde groep	Gestandaardiseerd Residu	Cook's Distance	Verwijderd
22	1	0,125	0	2,644	, 2841	Nee
52	1	0,139	0	2,493	,2954	Nee
56	1	0,119	0	2,723	,2801	Nee
90	0	0,937	1	-3,870	1,0851	Ja
91	0	0,877	1	-2,666	,4982	Nee

Tabel 3b: eigenschappen opvallende casussen

Casus		DG	Intern of Extern	Besteding Publieke Middelen	Politieke Gevoeligheid Wetgeving	Evaluan-dum	Tijdsdruk
22	**Catalysts for Change? Equality bodies according to Directive 2000/43/EC - existence, independence and effectiveness**	Employment, social affairs and inclusion	Extern	Nee	79	Proces	Nee
52	**National Emission Ceilings Directive Review**	Environment	Extern	Nee	85	Proces	Nee
56	**Financial Security in Environmental Liability Directive, Final Report (Interim)**	Environment	Extern	Nee	76	Proces	Nee
90	**EU public procurement legislation: delivering results.**	Internal Market and Services	Intern	Nee	126	Oordeel	Ja
91	**Report on the Implementation and Effect of the Resale Right Directive**	Internal Market and Services	Intern	Nee	88	Oordeel	Nee

Tabel 3a laat zien dat de casus 22, 52 en 56 weliswaar afwijkend is, maar niet leidt tot een vertekening van het model, de *Cook distance* is met waarden onder de 0,3 ver verwijderd van het gevaarlijke niveau van 1. Het onderwerp van de evaluatie is de inrichting van een structuur die slachtoffers van raciale en etnische discriminatie moet bijstaan. Uit tabel 3b blijkt dat bij deze evaluatie, hoewel er geen sprake is van de besteding van publieke middelen of van het vellen van een oordeel, wel gekozen is voor een externe evaluatie. Wat voor een externe evaluatie spreekt, is het ontbreken van tijdsdruk en een ruim bovengemiddelde politieke gevoeligheid. De verklaring voor de grote standaard- afwijking lijkt dan ook in deze

politieke gevoeligheid te liggen.

Eenzelfde komt naar voren bij het bestuderen van de casussen 52 en 56. Casus 52 heeft het nakomen van de afspraken uit het National Emission Ceilings Directive om de uistoot van vervuilende stoffen te beperken tot onderwerp. Veel lidstaten lopen hierbij achter. Casus 56 richt zich op de transpositie van de richtlijnen die moeten zorgen dat milieuschade wordt verhaalt op de veroorzaker van deze schade tot onderwerp. Ook hier is sprake van achterstanden.

Casus 91 betreft een evaluatie van de richtlijn die moet zorgen voor een eenduidig systeem binnen de Europese Unie dat de rechten van auteurs beschermd bij economische exploitatie van hun werk. Hierbij worden geen definitieve conclusies getrokken. Deze casus heeft een grotere Cook's distance van bijna 0,5. Voor deze casus is gekozen voor een interne evaluatie. Er is dan ook geen sprake van besteding van publieke middelen. De evaluatie vindt echter niet onder tijdsdruk plaats, is politiek gevoeliger dan gemiddeld en is bedoeld voor het vellen van een oordeel. Het zou dan ook interessant zijn om diepgaander onderzoek te doen naar deze casus. Binnen de beperkte mogelijkheden van dit artikel is dit echter niet mogelijk. De Cook distance van ruim 0,5 laat zien dat dit ook niet noodzakelijk is. Hoewel deze waarde hoger is, dan bij de vorige casussen is ook dit nog ver verwijderd van het gevaarlijke niveau van 1.

Dit is anders bij casus 90. Onderwerp van deze evaluatie is de mate waarin verschillende richtlijnen hebben bijgedragen aan het creëren van transparantie en cross-border competitie bij de gunning van overheidsopdrachten. Dit betreft een intern onderzoek dat onder tijdsdruk heeft plaatsgevonden en waarbij geen sprake is van de besteding van publieke middelen. De politieke gevoeligheid is echter bovengemiddeld groot, wat niet verrassend is gezien het onderwerp. Ook heeft de evaluatie het geven van een oordeel tot doel. Dat deze casus een Cook distance van 1,0851 heeft, betekent dat deze casus een zo groot effect op het

model heeft dat de kans op vertekening erg groot is (Field, 2009, pp.292-293; Lammers e.a., 2007, p.162). Omdat het doel van dit artikel niet is om opvallende casussen diepgaand te onderzoeken, maar om een algemeen beeld te schetsen van de evaluatiepraktijk van de Europese Commissie wat betreft de keuze voor interne dan wel externe evaluatie is dan ook besloten deze casus uit de verdere analyse te verwijderen.

Nu de residuanalyse is uitgevoerd kan worden getest op multicollineariteit. Voordat de daadwerkelijke analyse zal worden uitgevoerd, zal eerst worden gekeken of er sprake is van (multi)collineariteit. Als sprake is van collineariteit hangen predictoren zo sterk samen dat de resultaten niet meer betrouwbaar zijn (Lammers e.a., 2007, p.160). Een bepaalde mate van collineariteit is echter nauwelijks te vermijden (Field, 2009, p.224). Om collineariteit te meten in een logistische regressie moet echter worden teruggegrepen naar de procedure die gebruikt wordt bij de 'normale' logistische regressie (Field, 2009, p.297). Een eerste mogelijkheid voor het meten van (multi)collineariteit biedt Spearman's rho. De procedure uit SPSS leidt tot de uitkomst in tabel 4 op de volgende pagina.

Tabel 4: Correlatie: Spearman's rho

		Besteding publieke middelen	**Politieke gevoeligheid wetgeving**	**Evaluatie gericht op een oordeel of op het proces**	**Tijdsdruk**
Besteding publieke middelen	*Spearman's Rho*				
	Sig. (2-tailed)				
	N				
Politieke gevoeligheid wetgeving	*Spearman's Rho*	-,071			
	Sig. (2-tailed)	,444			
	N	120			
Type evaluatie: oordeel of proces	*Spearman's Rho*	,266**	-,074		
	Sig. (2-tailed)	,001	,441		
	N	143	112		

Tijdsdruk	*Spearman's Rho*	,040	,164	,458**	
	Sig. (2-tailed)	,697	,143	,000	
	N	99	81	99	

**Correlation is significant at the 0.01 level (2-tailed).

De tabel laat zien dat bij een significantieniveau van 0,01 sprake is van collineariteit tussen twee groepen variabelen. Zo is er sprake van collineariteit tussen de variabelen 'Besteding van publieke middelen' en 'Type evaluatie: oordeel of proces'. Ook is er sprake van collineariteit tussen de variabelen 'Tijdsdruk' en 'Evaluatie gericht op een oordeel of op het proces'. Field (2009, p. 273) stelt echter dat collineariteit pas een probleem wordt zodra de correlatiecoëfficiënt de waarde van 0,8 overschrijdt. Op het eerste oog is er met waarden van respectievelijk 0,266 en 0,458 geen probleem. Om er zeker van te zijn dat er geen sprake is van collineariteit, zijn ook de *VIF-waarde* en de *Tolerance* gemeten. De resultaten van deze analyse zijn weergegeven in tabel 5.

Tabel 5: de VIF- & Tolerance waarden

	Tolerance	VIF
Besteding publieke middelen	,920	1,087
Type evaluatie: oordeel of proces	,757	1,321
Politieke gevoeligheid wetgeving	,928	1,078
tijdsdruk	,772	1,295

Voor het beoordelen van deze waarden zijn geen vaste regels, wel zijn er een aantal vuistregels (Field, 2009, p.224). Zo wordt algemeen aanvaard dat een *VIF-waarde* van boven de 10 reden is voor zorgen. Voor de verwante *Tolerance* geldt dat deze niet onder de 0,1 mag komen terwijl een waarde van 0,2 reden is tot zorgen. Tabel 5 bevestigt zo het beeld dat uit tabel 4 naar voren kwam. Er is wel sprake van enige collineariteit, maar niet zodanig er reden is tot zorgen. Immers, geen enkele variabele kent een *tolerance* van onder de 0,2 en/of een *VIF-waarde* van boven de 10.

Nu besloten is een casus te verwijderen moet de analyse met de methode *Backward LR* opnieuw worden uitgevoerd. Deze methode start met het opstellen van een model waarbij een voorspelling wordt gedaan op basis van alleen de constante. Dit getal representeert de passendheid van een model waarbij er vanuit wordt gegaan dat alle cases dezelfde, meest voorkomende uitkomst hebben. In dit geval extern. De -2 Log Likelihood waarde waarmee dit gemeten wordt, verbeeldt de afstand tussen de werkelijke waarde en de 'voorspelde waarde' van iedere case. De -2 Log Likelihood waarde van dit model is 102,961. Indien het model bij de data past, gaat deze waarde omlaag zodra een berekening wordt uitgevoerd op basis van het complete uit de literatuur afgeleide model met variabelen .

Dit is echter niet de enige methode om de passendheid van het model te testen. Ook de '*Hosmer and Lemeshow test'* is hiervoor een indicator (Lammers e.a., 2007, p.147; Starkweather, 2012, n.p.). Deze test geeft een χ^2-waarde van 5,190 met een bijbehorend significantie van 0,737. Dit betekent dat het model bij een significantieniveau van 0,05 geschikt is voor het voorspellen van de waargenomen data (Starkweather, 2012, n.p.). De kans dat de verschillen tussen geobserveerde en door het model voorspelde frequenties op toeval berusten is met een waarde van 0,737 groot (Lammers e.a., 2007, pp.147-148). Met het gebruik van deze test moet men echter voorzichtig zijn, zeker bij erg kleine steekproeven. De test zal in zo'n geval meestal aangeven dat het model past, ook wanneer dit niet het geval is. Bij heel grote steekproeven is juist de kans op een onterechte afkeuring van het model groot. De test geeft dan vaak aan dat er significante verschillen zijn, terwijl het model wel goed bij de data past (Sieben, 2002, n.p.).

De feitelijke logistische regressieanalyse met de methode Backward LR leidt tot de resultaten die in tabel 6 op de volgende pagina zijn weergegeven. De tabel laat zien dat de effecten voor alle variabelen bij een significantieniveau van 0,05 significant zijn.

Tabel 6: resultaten

Variabele	***B***	***SE***	***DF***	***Sig.***	***Exp (B)***
Besteding publieke middelen	*-2,706*	1,001	*1*	*0,007*	*0,067*
Type evaluatie: oordeel of proces	*-2,873*	,877	*1*	*0,001*	*0,057*
Politieke gevoeligheid wetgeving	,037	,013	*1*	*0,005*	*1,038*
Tijdsdruk	*-1,972*	1,007	*1*	*0,006*	*0,071*

Voor de eerste variabele ‘type evaluate: oordeel of proces’ werd de hypothese geformuleerd dat bij evaluaties waarbij het doel is te komen tot een beoordeling van succes of falen van wetgevingsprogramma de kans op een externe evaluatie toeneemt. De analyse wijst echter uit dat de kans op een externe evaluatie juist afneemt als sprake is als het doel is om het wetgevingprogramma te beoordelen. Tabel 6 laat zien dat het effect (‘B’) van een evaluatie met als doel het geven van een oordeel op de kans op een externe evaluatie -2,873 is. Ook de Odds Ratio (‘Exp (B)’) laat dit zien. De waarde van 0,057 betekent niet dat de kans op een externe evaluatie bij een evaluatie gericht op het vellen van een oordeel 0,057 maal zo groot is als bij een evaluatie gericht op het proces. Wel betekent dit dat de kans op een externe evaluatie veel kleiner is als het doel het beoordelen van de wetgeving is, dan wanneer het proces centraal staat (Field, 2009, pp. 270-271). De hypothese dat de kans op een externe evaluatie groter is bij een evaluatie met als doel het beoordelen van succes of falen van een evaluatie zal dan ook verworpen moeten worden.

De tweede hypothese ‘besteding publieke middelen’ stelde dat de kans op een externe evaluatie groter is indien de evaluatie betrekking heeft op wetgeving waarin de besteding van de publieke middelen expliciet aan de orde is gekomen. Ook hier laat de analyse een omgekeerd beeld zien. Bij wetgeving waar expliciet de besteding van publieke middelen aan de orde komt, is de kans op een externe evaluatie juist kleiner getuige de B-waarde van -2,706. Bij een evaluatie die betrekking heeft op wetgeving waarbij de besteding van publieke middelen expliciet aan de orde komt, neemt de kans op een externe evaluatie 2,706 af. De

Odds Ratio bevestigt met een waarde van 0,067 dat de kans op een externe evaluatie afneemt als de besteding van publieke middelen expliciet aan de orde wordt gesteld in de wetgeving. De hypothese dat de kans op een externe evaluatie groter is indien de evaluatie betrekking heeft op wetgeving waarin de besteding van de publieke middelen expliciet aan de orde is gekomen wordt dan ook verworpen.

Ten derde werd gesteld dat de kans op een externe evaluatie afneemt, indien een evaluatie onder tijdsdruk plaatsvindt. Deze verwachting wordt tijdens de analyse bevestigd zo laten de B-waarde van -1,972 en de Odds Ratio van 0,071 zien. De hypothese dat de kans op een externe evaluatie kleiner wordt, als de evaluatie onder tijdsdruk plaatsvindt, wordt dan ook niet verworpen en blijft gehandhaafd.

De laatste hypothese luidt dat de kans op een externe evaluatie toeneemt als de evaluatie betrekking heeft op politiek gevoelige wetgeving. De analyse laat zien dat bij een toename van de politieke gevoeligheid de kans op een externe evaluatie toeneemt met 0,037. Dit effect lijkt zeer beperkt in vergelijking met de andere drie variabelen. Bij de interpretatie van dit effect moet echter worden meegenomen dat de variabelen ‘Type evaluatie: oordeel of proces’, ‘Besteding publieke middelen’ en ‘Tijdsdruk’ een categorisch karakter hebben, terwijl de variabele ‘Politieke gevoeligheid wetgeving’ continue van aard is. De hypothese dat de kans op een externe evaluatie toeneemt als de evaluatie betrekking heeft op politiek gevoelige wetgeving, wordt dan ook niet verworpen.

Door te kijken naar de waarden in de tabel ‘Model if Term Removed’ is het mogelijk de relatieve importantie van de verschillende variabelen te bepalen (Lammers e.a., 2007, p.159). Deze waarden zijn in tabel 7 op de volgende pagina weergegeven.

Tabel 7: de relatieve importantie van de variabelen

Variabele	***Verandering in -2 Log Likelihood***	***Significantie***	***Relatieve Belang***
Besteding publieke middelen	-28,927	*,002*	*3*
Type evaluatie: oordeel of proces	-31,431,	*,000*	*1*
Politieke gevoeligheid wetgeving	-29,942	*,001*	*2*
Tijdsdruk	-28,683	*,003*	*4*

Ook uit deze tabel blijkt dat geen enkele variabele bij een significantieniveau van 0,05 kan worden verwijderd zonder dat het model hier onder lijdt. Ten tweede blijkt uit de tabel dat het effect van het verwijderen van de variabele 'Type evaluatie' het grootst is, een verwijdering van deze variabele zou de grootste gevolgen hebben voor de -2 Log Likelihood. Het relatieve belang van de variabele 'Tijdsdruk' is het kleinst.

Dat geen enkele variabele kan worden verwijderd, wordt ook weerspiegeld in de verandering die optreedt in de -2 Log Likelihood waarde. Deze daalt door de toevoeging van de variabelen sterk en heeft nu een waarde van 46,659 tegenover een waarde van 102,961 voor het model zonder de variabelen. Een model waarin de gesproken variabelen zijn opgenomen is dus veel sterker dan een model zonder deze variabelen. Om uitspraken te doen over de verklaringskracht van het model zijn echter nog andere gegevens nodig. Omdat er bij een dichotome variantie niet gesproken kan worden van variantie is er geen 'R Square'-waarde in de logistische regressieanalyse om de verklaarde significantie weer te geven (Sieben, 2002, n.p.). Er zijn andere maten die beschouwd kunnen worden als vervangers van deze maat. Dit geldt voor de *'Nagelkerke R Square'*, de *'Cox & Snell R Square'*(Field, 2009, p.269) en de *McFadden R Square* (Lammers e.a., 2007, p.153). Deze laatste wordt niet door SPSS gegeven, maar kan eenvoudig worden berekend door de -2 Log Likelihood van het

volledige model af te trekken van de -2 Log Likelihood van het basismodel en de uitkomst hiervan te delen door de -2 Log Likelihood van het basismodel. Ieder afzonderlijk zeggen deze maten niet zoveel, maar in combinatie krijgen deze hun betekenis (Lammer e.a., 2007, p.147). Voor alle drie de waardes geldt dat deze een waarde aannemen tussen 0 (geen verklaringskracht) en 1 (volledige verklaringskracht). Hierbij moet worden opgemerkt dat de Cox en Snell R Square nooit zijn theoretische maximum van 1 bereikt (Field, 2009, p. 269; Lammers e.a., 2007, p.144). In dit onderzoek heeft de Cox & Snell R Square een waarde van 0,506. De Nagelkerke R Square heeft zelfs een waarde van 0,686. De McFadden R Square bedraagt ([102,961- 46,659] / 102,961=) 0,547. Aan de hand van deze waardes kan worden geconcludeerd dat het model niet in staat is de keuze tussen interne en externe evaluatie volledig te verklaren, maar wel degelijk in staat is de keuze tussen intern en extern voor een aanzienlijk deel te verklaren: ongeveer de helft van de spreiding tussen interne en externe evaluatie kan met behulp van het model verklaard worden.

Conclusie

De positie van de Europese Commissie als vertegenwoordiger van het Europese niveau is na een crisis aan het einde van de jaren ’90 steeds sterker geworden. Centraal begrip in deze ontwikkeling is subsidiariteit. Hierbij is het van groot belang dat de Europese Commissie haar toegevoegde waarde ten opzichte van het nationale en het regionale bestuursniveau laat zien. Hiervoor maakt zij onder andere gebruik van ex-post evaluaties van haar activiteiten. Zij kan er voor kiezen deze evaluaties zelf uit te voeren of door externen uit te laten voeren. Een onderzoek naar deze keuze zal meer inzicht verschaffen in de rol van politieke factoren in de evaluatiepraktijk van de Commissie. Is bij de keuze tussen interne en externe evaluatie een rol weggelegd voor politieke factoren?

Voor de beantwoording van deze vraag is een model opgesteld om de keuze tussen interne en externe evaluatie in de Europese Commissie te verklaren. Uit dit model is gebleken

dat in tegenstelling tot de uit de gangbare praktijk van evaluatieonderzoek afgeleide verwachtingen de kans op een interne evaluatie door de Europese Commissie juist groter is als er publieke middelen zijn besteed. Ook als er een oordeel moet worden gegeven wordt de kans op een interne evaluatie, tegengesteld aan de verwachting, groter. De kans op een interne evaluatie wordt ook groter indien de evaluatie onder tijdsdruk moet plaatsvinden, dit is echter in lijn met de verwachting. De kans op een externe evaluatie wordt, overeenkomstig de hypothese, vergroot indien sprake is van een gevoelig beleidsterrein.

Deze resultaten zijn een duidelijke aanwijzing dat de evaluatiepraktijk van de Europese Commissie op het gebied van wetgevingsevaluaties afwijkt van de 'gangbare' theorieën over beleidsevaluatie zoals deze in de verschillende standaardwerken wordt gepresenteerd. De Commissie lijkt dan ook een heel eigen logica te hanteren voor het al dan niet inhuren van externen voor het evalueren van haar wetgevingsactiviteiten. Dit is niet direct in overeenstemming met de technocratische benadering die de Commissie hanteert als zij over haar evaluatiepraktijk spreekt. In plaats daarvan lijkt de vergrote kans op een interne evaluatie bij het beoordelen van wetgeving en de controle van bestede gelden te wijzen op een verlangen om zelf de regie in handen te houden. Dat evaluaties op een politiek gevoelig beleidsterrein wel vaker door derden gehouden worden, doet hier niets aan af. Pas bij een hoge politieke gevoeligheid neemt de kracht van dit argument toe. Een kritische benadering van de evaluatiepraktijk van de Europese Commissie lijkt dan ook op zijn plaats. Hiervoor is het aan te bevelen om de in dit onderzoek gesignaleerde opvallende evaluaties diepgaand te onderzoeken om zo een volgende stap te zetten naar meer kennis over de evaluatiepraktijk en het mogelijke politieke gebruik van evaluaties binnen de Europese Commissie. Ook een uitgebreider onderzoek naar politieke factoren binnen de evaluatiepraktijk van de Commissie is sterk aan te bevelen.

Literatuur

- Barosso, J.M. (2009). *Political guidelines for the next Commission.* Vinddatum 28 maart 2012 op *http://ec.europa.eu/commission_2010-2014/president/pdf/press_20090903_en.pdf*
- Bovens, M., 't Hart, P. & Kuipers, S. The Politics of Policy Evaluation. In *The Oxford Handbook of Public Policy (pp. 319-335)*[electronic version]. Vinddatum 2 april 2012 op: http://www.oxfordhandbooks.com.proxy.ubn.ru.nl:8080/oso/private/content/oho_politics/9780199548453/p031.html#oxfordhb-9780199548453-chapter-015
- Bressers, J.Th.A. & Hoogerwerf, A. (1991). Inleiding tot de beleidsevaluatie In Bressers, J.Th.A. & Hoogerwerf, A [Eds.] *Beleidsevaluatie* (pp. 19-32). Alphen aan den Rijn: Samsom H.D. Tjeenk Willink.
- Chryssochoou, D.N. (2010). Europe's Contested Democracy. In: Cini, M. & Perez-Solórzano Borragán, N. [Eds.] *European Union Politics [3rd Edition]*. Pp. 377-390. Oxford: Oxford University Press
- Cini, M. & Perez-Solórzano Borragán, N. (2010). Glossary. In: Cini, M. & Perez-Solórzano Borragán, N. [Eds.]. *European Union Politics [3rd Edition]*. Pp. 437-453. Oxford: Oxford University Press
- Conley-Tylor, M. (2005). A fundamental Choice: internal of external evaluation? In: *Evaluation Journal of Australia*, Vol. 4 ,Issue 1, pp. 3-11 [Electronic Version].
- Cummings, O.W., Nowakowski, J.R., Schwandt, T.A., Eichelberger, R.T., Kleist, K.C., Larson, C.L. & Knott, T.D. (1988). Business perspectives on internal/external evaluation [Electronische Versie]. In: *New Directions for Program Evaluation,* Vol. 1988, Issue 39 pp. 59–74.
- Curtin, D. & Egeberg, M. (2008). Tradition and Innovation: Europe's Accumulated Executive Order. *West European Politics,* Vol. 31, Issue. 4 [Electronic Version]. Vinddatum 26 april 2012 op http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1349784#captchaSection
- Deetman, W. (1988). Macht en onmacht van evaluatie van en door bestuur. In: Ringeling, A.B. & Sorber, A. [Eds.] *Macht en onmacht van bestuurlijke evaluaties* (pp. 27-37). 's Gravenhage: Vuga.

- Derlien, H.U. & Rist, R.C. (2002). Conclusion: policy evaluation in International Comparison. In: Furubo, Rist & Sandahl [Eds.] *International Atlas of Evaluation* (pp. 439-455). London: Transaction Publishers.
- Directorate General for the Budget (n.d.) *Evaluation in the Commission; Reporting on Results. Annual Evaluation Overview 2000: Conclusions and findings from evaluations in the Commission*
  Vinddatum 11 mei 2012 op
  http://ec.europa.eu/dgs/secretariat_general/evaluation/documents_en.htm
- Directorate General for the Budget (n.d.) *Evaluation in the Commission; Reporting on Results. Annual Evaluation Overview 2001: Conclusions and findings from evaluations in the Commission*
  Vinddatum 11 mei 2012 op
  http://ec.europa.eu/dgs/secretariat_general/evaluation/documents_en.htm
- Directorate General for the Budget (n.d.) *Evaluation in the Commission; Reporting on Results. Annual Evaluation Overview 2002: Conclusions and findings from evaluations in the Commission*
  Vinddatum 11 mei 2012 op
  http://ec.europa.eu/dgs/secretariat_general/evaluation/documents_en.htm
- Directorate General for the Budget (n.d.) *Evaluation in the Commission; Reporting on Results. Annual Evaluation Overview 2003: Conclusions and findings from evaluations in the Commission*
  Vinddatum 11 mei 2012 op
  http://ec.europa.eu/dgs/secretariat_general/evaluation/documents_en.htm
- Directorate General for the Budget (n.d.) *Evaluation in the Commission; Reporting on Results. Annual Evaluation Overview 2004: Conclusions and findings from evaluations in the Commission*
  Vinddatum 11 mei 2012 op
  http://ec.europa.eu/dgs/secretariat_general/evaluation/documents_en.htm
- Directorate General for the Budget (n.d.) *Evaluation in the Commission; Reporting on Results. Annual Evaluation Overview 2005: Conclusions and findings from evaluations in the Commission*

Vinddatum 11 mei 2012 op
http://ec.europa.eu/dgs/secretariat_general/evaluation/documents_en.htm

- Directorate General for the Budget (n.d.) *Evaluation in the Commission; Reporting on Results. Annual Evaluation Overview 2006: Conclusions and findings from evaluations in the Commission*
  Vinddatum 11 mei 2012 op
  http://ec.europa.eu/dgs/secretariat_general/evaluation/documents_en.htm
- Directorate General for the Budget (n.d.) *Evaluation in the Commission; Reporting on Results. Annual Evaluation Overview 2007: Conclusions and findings from evaluations in the Commission*
  Vinddatum 11 mei 2012 op
  http://ec.europa.eu/dgs/secretariat_general/evaluation/documents_en.htm
- Directorate General for the Budget (n.d.) *Evaluation in the Commission; Reporting on Results. Annual Evaluation Overview 2008: Conclusions and findings from evaluations in the Commission*
  Vinddatum 11 mei 2012 op
  http://ec.europa.eu/dgs/secretariat_general/evaluation/documents_en.htm
- Directorate General for the Budget (n.d.) *Evaluation in the Commission; Reporting on Results. Annual Evaluation Overview 2009: Conclusions and findings from evaluations in the Commission*
  Vinddatum 11 mei 2012 op
  http://ec.europa.eu/dgs/secretariat_general/evaluation/documents_en.htm
- Duina, F. (1997). Explaining Legal Implementation in the European Union. *International Journal of the Sociology of Law.* Vol. 25, pp.155-179.
- Egeberg, M. (2010). The European Commission. In: Cini, M. & Perez-Solórzano Borragán, N. [Eds.] *European Union Politics [3rd Edition]*. Pp. 125-140. Oxford: Oxford University Press
- Euréval (2012). *Our project.* Vinddatum 10 oktober 2012 op http://www.eureval.org/-Our-project-.html

- Europese Commissie (2004). Evaluating EU activities: A practical guide for the Commission services. Vinddatum 31 mei 2012, op http://ec.europa.eu/dgs/secretariat_general/evaluation/docs/eval_activities_en.pdf
- Europese Commissie (2007). *Responding to Strategic Needs: Reinforcing the use of evaluation* [electronic version]. Vinddatum 6 maart 2012 op http://ec.europa.eu/dgs/secretariat_general/evaluation/docs/eval_comm_sec_2007_213_en.pdf
- Europese Commissie (2010). *Smart Regulation in the European Union* [Electronic version]. Vinddatum 8 mei 2012 op http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2010:0543:FIN:EN:PDF
- Europese Commissie; DG Enterprise & Industry. (2012). *Smart regulation: impact assessment.* Vinddatum 15 mei 2012 op http://ec.europa.eu/enterprise/policies/smart-regulation/impact-assessment/index_en.htm.
- Field, A. (2009). *Discovering statistics: using SPSS (and sex and drugs & rock 'n roll)* [3e edition]. SAGE: Londen.
- Formez (2004). A comparative analysis of RIA in ten EU countries. Report prepared for the Directors of Better Regulation, Vinddatum 22 juli 2012 op: https://www.wbginvestmentclimate.org/uploads/9.ReportonRIAEU.pdf
- Van Gerven, W. (2000) ' Ethical and Political Responsibility of EU Commissioners'. In: *Common Market Law Review*, Vol. 37, Issue 1, pp. 1–6.
- Herweijer, M. & Kotmann, R.H.P.W. (1988). Intern of Extern evaluatie-onderzoek? In Ringeling, A.B. & Sorber, A. [Eds.] *Macht en onmacht van bestuurlijke evaluaties* (pp. 91-104). 's Gravenhage: Vuga.
- Hofmann, H.C.H. (2008). Mapping the European Administrative Space. In D. Curtin & M. Egeberg [Eds.], *Towards a new executive order in Europe?* (pp. 24-38). London: Routledge Taylor & Francis Group.
- House, E. R. (1986). Internal Evaluation. In: *Evaluation Practice,* Vol. 7, Issue. 1, pp. 63-64 [Electronic Version].
- Institute for European Environmental Policy (2004*). Sustainable development in the European Commission's integrated impact assessment for 2003. Final report*. Brussels: European Commission.

- Kapsis, I. (2010). The Courts of the European Union (3rd Edition). In M. Cini & N. Pérez-Solórzano Borragán (Eds.), *European Union Politics* (pp.176-188). Oxford: Oxford University Press.
- Kelemen, R.D. (2006). Suing for Europe: Adversarial Legalism and European Governance. In: *Comparative Political Studies*, Vol. 39, Issue 1, pp. 101-127
- Klein Haarhuis, C. M. & Niemeijer, E. (2008). *Wet en werkelijkheid* [Electronic Version]. Den Haag: WODC. Vinddatum 16 mei 2012 op www.wodc.nl/images/ob267_volledige%20tekst_tcm44-105133.pdfKirkpatrick, C. & Lee, N. (2004). *A Pilot Study of the Quality of European Commission Extended Impact Assessments*. IARC Working *Paper Series No. 8. Manchester*: Impact Assessment Research Centre: University of Manchester. Vinddatum 22 juli 2012 op: at http://ageconsearch.umn.edu/bitstream/30580/1/ia040008.pdf
- Lammers, J., Pelzer, B., Hendrickx, J. & Eisinga, R. (2007). *Categorische data analyse met SPSS: Inleiding in de loglineaire analysetechnieken.* Assen: Van Gorcum.
- Lodge, M. (2008). Regulation, the regulatory state and European politics. *West European Politics,* Vol. 31, (1-2), pp. 280-301.
- Love, A.J. (1983). Developing effective internal evaluation. In: Woolridge, R.J. & House, E.R. [Editors in Chief] *New directions for program evaluation* (Vol. 20).
- Macmullen, A. (1999) 'Political Responsibility for the Administration of Europe: The Commission's Resignation March 1999'. In: *Parliamentary Affairs*, Vol. 52, Issue. 4, pp. 703–18.
- Madaus, G.F., Stufflebeam, D. & Scriven, M.S. (1983). Program Evaluation: A Historical Overview. In Madaus, G.F., Stufflebeam, D. & Scriven, M.S.[Eds]. *Evaluation Models* (pp. 3-22). Dordrecht: Kluwer-Nijhoff Publishing.
- Majone, G. (1998). *Europe's 'Democratic Deficit': The Question of Standards.* In: *European Law Journal.* Vol.4 No. 1, pp. 5-28. [Electronic Version]
- Mathison, S. (1994). Rethinking the evaluator role: partnerships between organizations and evaluators. In: *Evaluation and Program Planning*, Vol. 17, Issue 3, pp. 299–304 [Electronic Version].

- Mayne, J., Bemelmans-Videc, M.L., Hudson, J. & Conner, R. [Eds.] (1992). *Advancing Public Policy Evaluation: learning from International Experiences.* Amsterdam: Elsevier Science Publishers B.V.
- McClave,J.T., Benson, P.G. & Sincich, T. (2007). *Statistiek: een inleiding voor het hoger onderwijs.* Amsterdam: Pearson Education Benelux
- Meuwese, A.C.M. (2008). Impact Assessment in EU Lawmaking. Vinddatum 28 maart 2012 op https://openaccess.leidenuniv.nl/bitstream/handle/1887/12589/Thesis.pdf?sequence=3
- Nugent, N. (2003). *The government and politics of the European Union (5th Edition).* New York: Palgrave Macmillan
- Opoku, C. and Jordan, A. (2004). *Impact Assessment in the EU: A Global Sustainable Development Perspective*. Vinddatum 22 juli 2012 op http://userpage.fu-berlin.de/ffu/akumwelt/bc2004/download/opoku_jordan_f.pdf
- Owen, J.M. (2007). *Program evaluation: Forms and Approaches* [3rd Edition]. New York: The Guilford Press.
- Papadopoulos, Y. (2010). Accountability and Multi-level Governance: More Accountability, Less Democracy? In: *West European Politics.* Vol. 33. No. 5. pp. 1030-1049 [Electronic Version]
- Princen, S. (2010). De institutionele architectuur van de EU. In: Van der Vleuten, A. (Ed.), *De bestuurlijke kaart van de Europese Unie*. *Instellingen, besluitvorming en beleid* (2e dr.). Bussum: Coutinho.
- Radaelli, C.M. (2005). Diffusion without convergence: how political context shapes the adoption of regulatory impact assessment. In: *Journal of European Public Policy*. Vol. 12, Issue 5, 924–43.
- Radaelli, C.M. (2007). Whither better regulation for the Lisbon agenda? In: *Journal of European Public Policy*. Vol. 14, Issue 2, pp.190-207.
- Radaelli, C.M. & Meuwese, A.C.M. (2010). Hard Questions, Hard Solutions: Proceduralisation through Impact Assessment in the EU. In: *West European Politics*. Vol.33, Issue 1, pp.136-153.

- Renda, A. (2006). *Impact Assessment in the EU: the state of the art and the art of the state*. Brussels: Centre for European Policy Studies
- Rosamond, B. (2010). New Theories of European Integration. In: Cini, M. & Perez-Solórzano Borragán, N. [Eds.] *European Union Politics [3rd Edition]*. Pp. 104-122. Oxford: Oxford University Press
- Rossi, P.H., Freeman, H.E. & Lipsey, M.W. (1999). *Evaluation: A systematic approach (6th Edition).* London: Sage Publications
- Sieben, I. (2002). *Logistische regressive analyse: een handleiding* [Electronic Version]. Vinddatum 30 april 2012 op http://www.ru.nl/socialewetenschappen/rtog/tips/onderdelen/logistische/
- Stame, N. (2008). The European Project, Federalism and Evaluation. In: *Evaluation.* Vol.14, Issue 2, pp.117-140.
- Starkweather, J. (2012). *SPSS Short Course: Binary Logistic Regression* [Electronic Version]. Vinddatum 5 september 2012 op http://www.unt.edu/rss/class/Jon/SPSS_SC/Module9/M9_LogReg/SPSS_M9_LogReg.htm
- Summa, H. & Toulemonde, J. (2002). Evaluation in the European Union. In: Furebo, J.E., Rist, C.R. & Sandahl, R. [Eds.]. *International Atlas of Evaluation.* Pp. 407-425. New Brunswick: Transaction Publishers.
- Swanborn, P. G. (2007). *Evalueren* [2nd Edition]. Amsterdam: Boom Onderwijs.
- Thatcher, M. & Coen, D. (2009). Reshaping European Regulatory Space: an Evolutionary Analysis. In: Curtin, D. & Egeberg, M. [Eds.]. *Towards a new Executive order in Europe?* Pp. 168-198. London: Routledge Taylor & Francis Group
- Toulemonde, J. (1996). Can we evaluate subsidiarity? Elements of answers from the European practice. In: *International Review of Administrative Sciences,* Vol. 62 Issue 1, pp. 49-61 [Electronic version].
- Tsakatika, M. (2005). Claims to Legitimacy: The European Commission between Continuity and Change. In: *Journal of Common Market Studies,* Vol. 43, Issue 1, pp. 193-220 [Electronic Version].

- Vedung, E. (1997). *Public Policy and Program Evaluation.* New Brunswick: Transaction Publishers.
- Vennix, J. (2007). *Theorie en praktijk van empirisch onderzoek.* Essex: Pearson Education Limited.
- Voorst, S. van (2012). *Topstudies en topregels: Onderzoek naar de kwaliteit van Europese ex-post wetgevingsevaluaties.* Nijmegen: Radboud Universiteit Nijmegen
- Weiss, C.H. (1972). *Evaluation research: methods for assessing program effectiveness,* Englewood Cliffs: Prentice Hall.
- Wildavsky, A. (1972). The selfevaluating organization. In: *Public Administration Review,* Vol. 32, Issue 5 pp. 509-520 [Electronic Version].
- Wille, A. (2007). *Bridging the Gap: Political and Administrative Leadership in a Reinvented European Commission.* Vinddatum 31 mei 2012 op http://www.ipa.udel.edu/3tad/papers/workshop1/Wille.pdf
- Worthen, B.R., Sanders, J.R. & Fitzpatrick, J.L. (1997). *Program Evaluation: Alternative approaches and Practical Guidelines* [Electronic Version]. Vinddatum: 27 maart 2012 op https://web.utk.edu/~cdavis80/EP521/readings/Worthen1.pdf