

Panel van Advies

Tweede Advies

Januari 2013

Inhoudsopgave

1. Inleiding.....	3
1.1 Achtergronden van het Panel van Advies	3
1.2 Mandaat van het Panel van Advies.....	3
1.2.1 Achtergrond	3
1.2.2 Taken van het panel	3
1.2.3 Samenstelling van het panel	4
1.3 Werkwijze Panel van Advies bij opstellen van het Tweede Advies.....	4
1.4 Belangrijke begrippen in het Tweede Advies	4
1.4.1 Bruikbaarheid en gebruik van bestudeerde rapporten	4
1.4.2 Evaluatieprogrammering	5
1.4.3 De rol van de beleidstheorie in evaluaties.....	5
1.5 Elementen van het mandaat gedekt in Tweede Advies.....	5
2. Bruikbaarheid van bestudeerde IOB-rapporten	6
2.1 Helderheid doel van het onderzoek	6
2.2 Beleidstheoretische verankering	7
2.3 Methodiek en kwaliteitsbewaking.....	8
2.4 Kwaliteit weergave bevindingen	9
2.5 Relatie doelstellingen en conclusies	10
2.6 Handvatten voor stakeholders	11
3. Gebruik bestudeerde IOB evaluaties.....	11
3.1 Zichtbaarheid van de rapporten.....	11
3.2 Toegankelijkheid.....	12
3.3 Feitelijk gebruik	12
4. De evaluatieprogrammering.....	13
4.1 Waarom dit thema?.....	13
4.2 Hoe heeft het panel dit vraagstuk aangevat?.....	13
4.3 Bevindingen.....	13
4.3.1 Relevante literatuur.....	13
4.3.2 Het vergelijkend onderzoek	14
4.3.3 Kernthema's.....	15
4.3.4 Enkele vragen voor de toekomst.....	20
5. 'Theories of change' en Interventietheorieën in evaluaties	21
6. Slotbeschouwing.....	23
Literatuurlijst.....	26

1. Inleiding

1.1 Achtergronden van het Panel van Advies

Tijdens het algemeen overleg met de vaste commissie voor Buitenlandse Zaken van de Tweede Kamer op 29 mei en 22 oktober 2008 heeft de toenmalige minister voor Ontwikkelingssamenwerking toegezegd een raad van advies in te stellen die hem adviseert over de bruikbaarheid en het gebruik van evaluaties.¹² Op 19 februari 2009 stuurde de minister het mandaat van het Panel van Advies aan de Kamer en heeft hij de vragen over het mandaat van de vaste commissie beantwoord.³⁴ In zijn brief van 25 mei 2010 informeerde de minister van Buitenlandse Zaken de Tweede Kamer over de samenstelling van het Panel van Advies. Op 4 juli 2011 heeft het Panel van Advies haar Eerste Advies aangeboden aan de Staatssecretaris. In september 2011 is het Advies door de Staatssecretaris aangeboden aan de Tweede Kamer met de bekendmaking van het vierde lid van het Panel van Advies.

1.2 Mandaat van het Panel van Advies

1.2.1 Achtergrond

Evaluaties hebben een tweeledige functie in de beleidscyclus van ontwikkelingssamenwerking. Zij dienen ter verantwoording over doelmatigheid en doeltreffendheid van het beleid en de beleidsuitvoering. Daarnaast geven zij inzicht in al dan niet bereikte resultaten en de factoren die daarvoor (mede-)bepalend zijn geweest. Daarmee hebben evaluaties een leerfunctie en dragen ze bij aan de kennisinfrastructuur voor ontwikkelingssamenwerking. Verantwoorden en leren zijn complementair. Alle evaluaties dienen in beginsel beide functies te vervullen.

Om de leerfunctie van evaluaties te versterken heeft de toenmalige Minister voor Ontwikkelingssamenwerking de Tweede Kamer op 19 februari 2009 toegezegd een Panel van Advies (PvA) in het leven te roepen, bestaande uit een aantal gerenommeerde externe deskundigen. Dit panel heeft tot taak door middel van advisering aan de betrokken bewindspersoon voor Ontwikkelingssamenwerking de bruikbaarheid en het gebruik van evaluatiestudies voor beleid en praktijk te bevorderen. In de bijlage werd het mandaat van het panel nader gespecificeerd.

Het PvA kiest uit zijn midden een voorzitter en regelt zelf zijn werkzaamheden en frequentie van vergaderen. Het panel zal zijn adviezen in geschreven vorm uitbrengen. Deze adviezen zijn openbaar. Zowel de betrokken bewindspersoon als de ambtelijke leiding van Ontwikkelingssamenwerking als ook IOB kan het panel verzoeken de uitgebrachte adviezen mondeling toe te lichten. Het PvA bestaat uit vier externe deskundigen.⁵ Deze zijn voor een periode van drie jaar benoemd, waarbij een eenmalige verlenging van de benoeming mogelijk is. De leden van het panel zijn onafhankelijk. Lidmaatschap van het panel sluit de mogelijkheid uit opdrachten in het kader van de door IOB uit te voeren evaluaties aan te nemen. Het secretariaat van het panel wordt gevoerd door IOB. De leden van het panel ontvangen voor hun werkzaamheden een passende vergoeding.

1.2.2 Taken van het panel

Ten aanzien van de bruikbaarheid van evaluaties:

- het adviseren over de uitgangspunten van de evaluatieprogrammering op het terrein van ontwikkelingssamenwerking alsmede over de werkwijze van IOB;

1 Tweede Kamer, vergaderjaar 2007-2008, 31 200 V, nr. 136.

2 Tweede Kamer, vergaderjaar 2008-2009, 29 237, nr. 83.

3 Tweede Kamer, vergaderjaar 2008-2009, 29 237, nr. 85.

4 Tweede Kamer, vergaderjaar 2008-2009, 29 237, nr. 99.

5 Dhr. prof. dr. F.L. Leeuw; Mevr. prof. dr. A. Niehof; Dhr. dr. J. Vandemoortele; Dhr. Prof. dr. J.B. Opschoor.

- het achteraf beoordelen van de validiteit en betrouwbaarheid van evaluatierapporten en van de uitvoerbaarheid van aandachtspunten en/of aanbevelingen, en het doen van aanbevelingen ter verbetering;

Ten aanzien van het gebruik van evaluaties:

- het adviseren over de institutionele verankering van de evaluatiefunctie in de beleidscyclus;
- het adviseren over maatregelen waarmee het leren van evaluaties wordt gestimuleerd;
- monitoring van de follow-up van aanbevelingen en/of aandachtspunten uit evaluatierapporten en het, zo nodig, doen van aanbevelingen voor verbetering.

1.2.3 Samenstelling van het panel

In het expertpanel zullen de volgende onderscheiden en complementaire expertises vertegenwoordigd zijn:

1. kennis van het functioneren van overheidsorganisaties en in het bijzonder de betekenis van de evaluatiefunctie binnen de overheid;
2. wetenschappelijke kennis op het terrein van ontwikkelingssamenwerking, relevante onderzoeksactiviteiten en methoden en technieken van toetsend onderzoek.

1.3 Werkwijze Panel van Advies bij opstellen van het Tweede Advies

In de zomer van 2011 is het panel van start gegaan met het bepalen van de werkwijze en de agenda van het Tweede Advies. Kijkende naar het mandaat van het panel en naar de onderwerpen die werden gedekt in het Eerste Advies, werd bepaald dat de volgende onderwerpen aan bod komen in het Tweede Advies: a) de bruikbaarheid en het gebruik van (een selectie van) IOB rapporten; b) de evaluatieprogrammering; c) de rol van beleidstheorieën in evaluaties. In 2011/2012 is het panel zes keer bij elkaar gekomen. Er zijn gesprekken gevoerd met IOB medewerkers en leden van de Tweede Kamer (OS woordvoerders) en er is opdracht gegeven voor een vergelijkend onderzoek naar programmering aan een externe deskundige (Helmsing 2012).

1.4 Belangrijke elementen in het Tweede Advies

1.4.1 Bruikbaarheid en gebruik van bestudeerde rapporten

Voor dit advies hebben de leden van het Panel van Advies zes evaluaties en/of beleidsdoorlichtingen bestudeerd die door IOB in 2010, 2011 en begin 2012 zijn uitgebracht op het terrein van Ontwikkelings-samenwerking en twee gezamenlijke evaluaties. Er is gekozen om naar de rapporten te kijken die zijn vastgesteld in de periode 2010-2011; een selectie van rapporten vastgesteld in de voorafgaande periode werd in het Eerste Advies besproken. Gegeven de capaciteit van het PvA en het belang dat door de RPE wordt gehecht aan beleidsdoorlichtingen heeft het panel in het geval van beleidsdoorlichtingen er voor gekozen om de evaluatierapporten die aan de beleidsdoorlichting ten grondslag liggen niet in de studie te betrekken. De analyse van de IOB rapporten werd, afhankelijk van het type rapport, gedaan door een of twee panelleden per rapport die aan de hand van vooraf vastgestelde criteria de beoordeling uitvoerden. Indien er behoefte was aan nadere informatie op belangrijke punten, is er een toelichtend gesprek gevoerd met de betreffende IOB-inspecteur. Dit was het geval bij de evaluatie van capaciteitsontwikkeling (no. 336) en NiMD (no. 331).

Ter verdere bevordering van gebruik en doorwerking van de IOB-rapporten in beleid onderneemt IOB naast het verspreiden van de gedrukte exemplaren in binnen- en buitenland doorgaans het volgende: uitbrengen van een korte evaluatie nieuwsbrief met de hoofdbevindingen en aandachtspunten voor toekomstig beleid; publicatie van de rapporten op de website van het ministerie van Buitenlandse Zaken; feedbackbijeenkomsten voor BZ-medewerkers en externe belangstellenden; workshops in het betreffende ontwikkelingsland; presentatie van de evaluatie in internationaal verband (o.m. EU, OESO/DAC en op internationale bijeenkomsten van collega evaluatiediensten); melding van de evaluatie op sociale media (Twitter); en, waar relevant, verzoek tot melding van de evaluatie op de website van OESO/DAC en opname van het rapport in de *evaluation directory* van deze organisatie.

1.4.2 Evaluatieprogrammering

In dit advies wordt ook aandacht geschonken aan de uitgangspunten en de praktijk van de evaluatieprogrammering op het terrein van Ontwikkelingssamenwerking. Het PvA koos er voor om de Nederlandse praktijk en uitgangspunten van de evaluatieprogrammering in vergelijkend perspectief te plaatsen. Daartoe werd een externe deskundige, Prof. Dr. A.H.J. Helmsing (ISS), gevraagd om onderzoek te doen naar de manier waarop de evaluatieprogrammering tot stand komt en welke implicaties dit heeft bij de evaluatiediensten op het terrein van ontwikkelingssamenwerking van Zweden, Denemarken, en het Verenigd Koninkrijk. Daarnaast hebben de leden van het PvA interviews gehouden met een tweetal IOB-medewerkers en met leden van de vaste commissie van Buitenlandse Zaken van de Tweede Kamer, die een belangrijke afnemer zijn van IOB-rapporten. De belangrijkste resultaten van deze onderzoeksactiviteiten worden gepresenteerd in hoofdstuk 4.

1.4.3 De rol van de beleidstheorie in evaluaties

In dit advies wordt tevens aandacht geschonken aan de rol van beleids- of interventietheorieën bij IOB evaluaties. Hoewel niet expliciet onderdeel van het mandaat van het PvA, vindt het panel dat deze theorieën een belangrijke rol spelen in evaluaties en derhalve belicht moeten worden. De bevindingen op dit punt worden gerapporteerd in paragraaf 2.2.

1.5 Elementen van het mandaat gedekt in Tweede Advies

Op basis van de bevindingen van de leden van het PvA na bestudering van de IOB-rapporten, de uitkomsten van het onderzoek naar de uitgangspunten in de evaluatieprogrammering en de rol van de beleidstheorie in evaluaties worden de volgende punten conform het mandaat van het panel gedekt. Het advies biedt een analyse van de uitgangspunten, positionering en de praktijk van de evaluatieprogrammering op het terrein van ontwikkelingssamenwerking. Op basis van het achteraf beoordelen van de validiteit en betrouwbaarheid van evaluatierapporten van de rapporten beoogt het PvA bij te dragen aan de bevordering van de bruikbaarheid van de rapporten. Het gebruik van evaluaties wordt in dit advies mede belicht vanuit de invalshoek van enkele Tweede Kamerleden.

2. Bruikbaarheid van bestudeerde IOB-rapporten

Het 'groene boekje' van IOB, getiteld: 'Evaluatiebeleid en richtlijnen voor evaluaties' geeft aan dat de bruikbaarheid van evaluaties naar haar oordeel afhangt van de volgende factoren:⁶

- De helderheid van de specificatie van het (buiten het onderzoek zelf gelegen externe) doel van de evaluatie, waarvoor de onderzoeksuitkomsten kunnen worden gebruikt;
- De duidelijkheid en volledigheid waarmee in het evaluatierapport en de samenvatting van het rapport de essentie van het onderzoek en vooral de hoofdbevindingen zijn weergegeven;
- De volledigheid waarmee alle onderzoeksvragen door de conclusies worden beantwoord;
- De mate waarin de uitkomsten van het onderzoek aanleiding geven tot handelingen die binnen het vermogen van de ontvanger van het rapport liggen;
- De mate waarin de evaluatie antwoorden geeft op gestelde vragen.

Deze elementen van bruikbaarheid zijn verwerkt in de criteria die door het Panel zijn gebruikt.

De volgende rapporten zijn bestudeerd⁷:

- IOB, De Methodische kwaliteit van Programma-evaluaties in het Medefinancieringsstelsel-I 2007-2010, no. 362 (2011);
- IOB, Evaluatie van de Twinningfaciliteit Suriname-Nederland, no 361 (2011);
- IOB, Assisting Earthquake victims: Evaluation of Dutch Cooperating aid agencies (SHO) Support to Haiti in 2010, no. 358 (2011);
- IOB, Education matters: Policy review of the Dutch contribution to basic education 1999–2009, no. 353 (2011);
- Channel Research (2011) *Misschien Vrede – gezamenlijke evaluatie van conflictpreventie en vredesopbouw in de Democratische Republiek Congo, Syntheseverslag*. Brussel: FOD Buitenlandse Zaken, Buitenlandse Handel en Ontwikkelingssamenwerking.
- Bennett, J., S. Pantuliano, W. Fenton, A. Vaux, C. Barnett, and Brusset, E. (2010) *Aiding the Peace. A Multi-Donor Evaluation of Support to Conflict Prevention and Peacebuilding Activities in Southern Sudan 2005 – 2010*. ITAD Ltd., United Kingdom.
- IOB, Facilitating Resourcefulness. Synthesis report of the Evaluation of Dutch support to Capacity Development, no. 336 (2011);
- IOB, Evaluation of Dutch support to Capacity Development. The case of the Netherlands Institute for Multiparty Democracy (NIMD), no. 331 (2011).

Het PvA heeft met deze selectie getracht zo veel mogelijk recht te doen aan de diversiteit van de IOB rapporten.⁸ Zo is een beleidsdoorlichting onderdeel van het pakket (opgebouwd uit een literatuurstudie en impactevaluaties in diverse landen), een metastudie, 'theory-driven' evaluaties en wat wel genoemd worden, 'evidence-based' evaluaties, maar ook gezamenlijke evaluaties met andere donoren. De rapporten zijn bekeken aan de hand van criteria die mede zijn gebaseerd op de (methodologische) criteria genoemd in het 'groene boekje'.

2.1 Helderheid van doel en vraagstelling van het onderzoek

Ten aanzien van de evaluatie van de SHO-noodhulp aan Haïti kan worden gesteld dat de vragen helder zijn en de ToR zeer gedetailleerd. De wijze waarop de vele vragen in tabel 4 van de ToR werden gereduceerd tot de 14 onderzoeksvragen is minder duidelijk uitgelegd. De conceptuele inkadering van de vragen en hun relatering van aan de internationale afspraken over en definiëring van de relevante aspecten van noodhulp is zorgvuldig.

6 IOB, 'Evaluatiebeleid en richtlijnen voor evaluaties' 2009, p. 45.

7 Omdat de Beleidsdoorlichting Water en Sanitatie later uitkwam dan oorspronkelijk gepland, kon dit rapport niet worden meegenomen in dit advies.

8 Studies, systematic reviews en evalueerbaarheidsstudies zijn in de assessment periode van dit Advies niet uitgekomen.

Daarmee wordt het concept noodhulp, ondanks de inherente complexiteit ervan, concreter, iets wat de bruikbaarheid van het rapport ten goede komt.

De evaluatie van de Twinningfaciliteit Suriname-Nederland heeft een heldere en duidelijk afgebakende vraagstelling, namelijk of de twinning tussen de Surinaamse en Nederlandse organisaties een effectief instrument is om het maatschappelijk middenveld in Suriname te versterken. Of twinning bijdraagt aan armoedebestrijding is als zodanig niet opgenomen omdat de impact op armoedevermindering niet kon worden gemeten. Daarbij stond het ook niet expliciet opgenomen in het beleidskader waardoor het formeel geen punt van evaluatie was. Desondanks heeft IOB een poging gedaan om er iets over te zeggen wat vanuit het oogpunt van de beleidsrelevantie van het onderzoek is toe te juichen.

In het geval van het Sudan onderzoek lijken de hoofddoelen voldoende helder gesteld, ook al is voor sommige gehanteerde begrippen nadere specificatie wenselijk (bijvoorbeeld 'peace dividend'). Voor wat betreft de andere studies was er voor het Panel geen reden tot met maken van opmerkingen.

2.2 Beleidstheoretische verankering

De beleidstheoretische verankering van de geëvalueerde interventies komt volgens de rapporten in de verschillende interventies op uiteenlopende wijzen aan bod.

De joint evaluatie over Zuid-Sudan is helder over de 'overheersende' veranderingstheorie in 2005 maar constateert ook dat daar gebreken aan kleefden: in de interventies ging men ervan uit dat de conflicten voorbij waren en dat daarom aan 'recovery' kon worden begonnen en aan economische ontwikkeling, terwijl er in het gebied in feite nog tal van conflicten speelden. Deze diagnose is in het rapport geadstrueerd met reconstructies van conflictanalyses op verschillende momenten, zoals die op de betreffende momenten gemaakt hadden kunnen worden op basis van toen beschikbaar materiaal.

Het rapport over de SHO-noodhulp aan Haïti maakt gebruik van de OECD-DAC criteria voor de evaluatie van humanitaire hulp, het *Sphere Handbook* uit 2004 en de *Code of Conduct for the International Red Cross and Red Crescent Movement and Non-Governmental Organisations (NGOs) in Disaster Relief*. Hoewel het hier strikt genomen niet om een beleidstheorie gaat, zijn de operationele criteria in genoemde bronnen gebaseerd op het uitgangspunt dat hulp als deze vorm krijgt vanuit een humanitair imperatief.

In de metastudie over de kwaliteit van de MFS-I evaluaties is de verankering goed aangegeven: de studie werkt binnen het kader van enkele elementen uit het DAC-raamwerk het begrip capaciteitsontwikkeling uit in termen van 'resourcefulness' (à la Kaplan), waarbij de subjecten van de ontwikkelingsactiviteiten worden gezien als opererend in open systemen en capaciteit wordt gedefinieerd in termen van een aantal specifieke 'capabilities'.

In het geval van de evaluatie van het NiMD maakt IOB opmerkingen over de kwaliteit of de afwezigheid van een 'theory of change'. IOB beschouwt de aanwezigheid van een dergelijke theorie daar kennelijk als evaluatiecriterium. In dat geval van NiMD vraagt het panel zich af of deze constatering van de afwezigheid van een theory of change terecht is. Er is immers een NiMD-handboek uit 2004 dat wel degelijk ingaat op de samenhang tussen democratie en ontwikkeling. In de metastudie over de kwaliteit van MFS-I evaluaties wordt geconstateerd dat het ook veelal 'echt mis' gaat bij de interventielogica van de MFS-organisaties. Ook bij de gezamenlijke evaluatie over de conflictpreventie en vredesopbouw in de DRC komt de *theory of change* aan bod. De evaluatiestudie verwerpt de bruikbaarheid daarvan in de onderhavige situatie, wegens de diversificeringen en fluïditeiten in de concrete situatie, en opteert in plaats daarvan voor een conflict- en conflict gevoeligheidsanalyse waarbij niet (zoals in een veranderingstheorie) wordt uitgegaan van de doelstellingen van de interventies, maar van de onderscheiden 'conflict- of vredesfactoren'.

De vraag die bij het panel ontstond is of *theory of change* een bruikbaar onderdeel is van evaluatiekaders. Waar, wanneer en of moet aan een dergelijke theorie aandacht worden besteed, resp. worden 'ingezet'? Is het vervolgens terecht om (harde) conclusies te trekken over gebrek aan evidentie, indien de beleids- dan wel veranderingstheorie ontbreekt of (inhoudelijk) zwak of matig is? Op deze vragen wordt dieper ingegaan in hoofdstuk 5.

2.3 Methodiek en kwaliteitsbewaking

Over het algemeen geven alle rapporten inzicht in de toegepaste methodiek waardoor het voor het panel mogelijk was om te bekijken of deze adequaat was voor de vraagstelling. De twee gezamenlijke evaluaties ('Congo' en 'Sudan') hebben een veelheid aan relevante literatuur verzameld en op een adequate manier aangewend. Vanwege de afwezigheid van een goede en systematisch verrichte achtergrondstudie en monitoring, waren er in het geval van de Sudan-evaluatie weinig empirische data vooraf beschikbaar. Dit werd opgevangen door uit interviews opinies en inschattingen te destilleren – iets dat in zulke omstandigheden een werkbare oplossing is.

Ook de methodiek voor de beoordeling van de kwaliteit van MFS-I evaluaties is in orde. Echter het rapport is onduidelijk over enkele basale methodologische aspecten van documentenanalyse. Er wordt geen verwijzing gegeven naar literatuur over de beoordeling van evaluaties, zoals dat het geval is bij 'systematic reviews' of andere typen overzichtsstudies (Gough et al, 2011) die minder stringent vasthouden aan de 'golden standard' (van het [quasi-] experiment). Daarbij is ten eerste te denken aan overzichtsstudies waarin de 'realist evaluation' aanpak wordt gevolgd, die naar configuraties van *Contexts, Mechanisms & Outcomes* verwijst (Pawson, 2006) én daarnaast naar overzichtsstudies die *systematic reviews* én *realist synthesis* werk combineren (Van der Knaap et al, 2008).

Ten aanzien van de studie inzake de synthese capaciteitsontwikkeling wordt geconstateerd dat in geringe mate gebruik is gemaakt van literatuur uit andere disciplines. Leidend was een ECDPM studie (gelinkt aan de studie van Senge, 1990, handelend over 'systeem denken') naar de vijf *core capabilities*. Het werk van McKinsey & company (2001) en McMahon (2008) zijn gebruikt om de ECDPM studie in perspectief te plaatsen. Literatuur over perverse effecten van capaciteitsontwikkeling en 'goal displacement' wordt niet genoemd. Er is wel een uitgebreid hoofdstuk gebaseerd op literatuur over de setting van capaciteitsontwikkeling en donorsteun. De combinatie van deze beschouwing met de empirische bevindingen uit de 28 cases leidt tot een analyse over de effectiviteit van de Nederlandse steun aan capaciteitsontwikkeling.

De synthesestudie capaciteitsontwikkeling is gebaseerd op opvattingen van Kaplan (1999, 2000), voor een studie met een leerfunctie een nuttig uitgangspunt en hanteert als begrippenkader een systeem van 5 zogenaamde 'core capabilities'⁹. Er is ook gebruik gemaakt van een 'results chain' waarin verschillende actoren/ontvangers met elkaar zijn verbonden door resultaten op diverse niveaus die de actoren horizontaal verbinden. Daarin is 'outcome' (en zelfs impact) het ultieme resultaat. Bij hantering hiervan bleek men in de case studies en op basis van de informatie van de zuidelijke organisaties als regel niet veel verder te komen dan een inventarisatie van outputs. De vraag is bovendien of het hanteren van een lineaire resultaatketen wel goed mogelijk is in een situatie die wordt beschreven met een open systeemmodel en de focus daarin op een meer circulaire endogeniteit (ontwikkeling van binnenuit in plaats van gedreven door exogene factoren) en complexiteit.

In verschillende IOB evaluaties kunnen impact en duurzaamheid niet worden getoetst vanwege het beperkte tijds kader van de interventie in relatie tot de effectmeting. Een voorbeeld hiervan is de evaluatie Twinning-

⁹ Baser, H. en Morgan, P. (2008) Capacity, Change and Performance: study report. Maastricht: ECDPM.

faciliteit Nederland/ Suriname. Het programma veronderstelt dat: 1) versterkte Surinaamse organisaties leiden tot meer kritische massa op het maatschappelijk middenveld in Suriname; en 2) dat (mede) daardoor wordt bijgedragen aan armoedevermindering in Suriname (ook al staat de bijdrage aan armoedebestrijding niet expliciet in het beleidskader). Zoals gezegd in paragraaf 2.1, viel de tweede hypothese buiten de vraagstelling omdat de impact op armoede geen onderzoeksvraag was. Wel heeft IOB een poging gedaan om daarvan een inschatting te maken, gebruikmakend van de drie OECD-DAC componenten: financiële armoede, gebrek aan toegang tot publieke basisdiensten en gebrek aan 'voice'.

Voor de *kwaliteitsbewaking* hanteert IOB een referentiegroep, voorgezeten door de leiding van IOB, die commentaar (advies) levert op de kwaliteit van de ToR en het eindrapport. De referentiegroep bestaat uit een of enkele externe leden en stakeholders van het beleid (beleidsmedewerker van het ministerie of NGO). Het advies van de referentiegroep kan worden verwerkt in het rapport. De directeur IOB stelt het evaluatierapport uiteindelijk vast. In alle gevallen heeft dit proces plaatsgevonden; in een enkel geval niet zonder een kritische kanttekening van een externe betrokkene (synthesestudie capaciteitsontwikkeling). De vormgeving van de kwaliteitsbewaking voor gezamenlijke evaluaties wijkt enigszins af van evaluaties die volledig onder verantwoordelijkheid van IOB vallen. De gezamenlijke evaluatie over de steun aan conflict preventie en vredesopbouw in Zuid-Sudan werkte met een gelaagde structuur met o.a. een Evaluation Managementgroep waarin IOB zitting had en die toezag op de kwaliteit van het rapport, en drie onafhankelijke wetenschappers voor externe kwaliteitsborging. In het geval van de Twinning faciliteit Suriname-Nederland was er een referentiegroep in Nederland en in Suriname om het rapport te bespreken.

Voor de synthesestudie capaciteitsontwikkeling is de opzet van de werkwijze anders dan de standaard IOB evaluatie. Er is sprake geweest van drie evaluaties die onder verantwoordelijkheid van IOB zijn uitgevoerd en vier onder verantwoordelijkheid van deelnemende maatschappelijke organisaties. Wat betreft laatstgenoemde evaluaties kan worden gesteld dat de 'onderzoeksstijl' en 'benaderingsstijl' van de consultant bepalend waren voor de kwaliteit van het rapport, ongeacht de coördinerende rol die IOB heeft gespeeld. Het syntheserapport is onder verantwoordelijkheid van IOB geproduceerd.

IOB betreft het evaluatieobject doorgaans bij de evaluaties door overleg, zoals tussentijdse updates of overleg tijdens referentiegroep bijeenkomsten. Het PVA plaatst een kanttekening bij de evaluatie van de methodische kwaliteit van programma-evaluaties van MFS-I organisaties.

In het geval van capaciteitsontwikkeling heeft IOB het evaluatieobject vastgesteld in overleg met Nederlandse en zuidelijke organisaties. Om het zuidelijk perspectief te waarborgen in het analytische kader heeft er een lokale uitwerking van het *5C-framework* plaatsgevonden (lokale kalibratie). Deze benadering staat wel op gespannen voet met de wens van generaliseerbaarheid. Fundamenteel is dat lokale kalibratie het gevaar van immunisering in zich draagt. Immers, er kán sprake zijn geweest van een selectie van bevindingen die goed aansluiten bij, respectievelijk passen bij de 5C-benadering terwijl 'lokale weerleggingen' daarvan niet onderkend of meegenomen zijn. Het gevolg is dan dat aangegeven wordt dat het *framework* 'paste', 'voldeed' of in vergelijkbare termen wordt gesproken, maar dat niet-passende ontwikkelingen buiten het frame gehouden kunnen zijn. Een dergelijke bias is meermalen in evaluaties in het algemeen aangetoond.

2.4 Kwaliteit weergave bevindingen

In vrijwel alle rapporten worden naar het oordeel van het Panel de bevindingen inzichtelijk weergegeven. In het geval van de DRC evaluatie wordt het beperkt aantal aanbevelingen als nuttig en doelgericht (op donoren en/of de uitvoerende organisaties) beschouwd. Ze volgen logischerwijs uit de gemaakte analyses. Hetzelfde wordt geconstateerd voor het merendeel van de bekeken rapporten. De bevinding voor MFS-I zijn technisch zeer inzichtelijk weergegeven. Het is alleen onduidelijk wat er achter de gegeven algemene oordelen schuilt aan specifieke observaties, wat het vermoedelijke leereffect niet ten goede komt. Verder wordt er in het

rapport terecht op gewezen dat de kwaliteit van evaluatieonderzoek moet worden onderscheiden van de kwaliteit van de geëvalueerde activiteiten en dat er op grond van de kwaliteit van de programmaevaluaties geen conclusies getrokken kunnen worden over de doelmatigheid en doeltreffendheid van de programma's zelf. Het PVA vermoedt dat de consequenties van dit standpunt onvoldoende doordacht zijn. Volgens Rossi's *Stainless Steel Law of Evaluation* is de kans kleiner dat er positieve effecten worden gevonden naarmate het evaluatiedesign bij impact studies strenger is. Meermalen is aangetoond dat de keuze voor een bepaald design van een evaluatie (experimenteel, quasi-experimenteel, longitudinaal, cross-sectioneel, case studie; vgl. De Vaus, 2001) een aanzienlijke invloed heeft op de uitkomsten.

Ondanks de hantering van een algemene ToR voor de synthesestudie over capaciteitsontwikkeling met een analytisch raamwerk voor zeven evaluaties en een kwaliteitsbewakingsmechanisme, heeft het IOB-evaluatieteam te maken gehad met weinig betrouwbare informatie over de behaalde resultaten op output en outcome niveau van de zuidelijke organisaties. Gebaseerd op matig basis materiaal (inclusief het repareren van tekortkomingen) is trekken van heldere en relevante conclusies een uitdaging geweest.

De evaluatie van de noodhulp aan Haïti heeft bij het opschrijven van de bevindingen terechte aandacht geschonken aan de complexe context (urbane regio, fragiele staat, veelheid aan donoren, etc.) waarin de hulp plaats vond.

2.5 Relatie doelstellingen en conclusies

De evaluaties van de kwaliteit van MFS-I evaluaties en de SHO-noodhulp aan Haïti waren helder in het toelichten van de relatie tussen doelstelling van de evaluatie, het beantwoorden van de vragen en de conclusies getrokken op basis van de empirische gegevens. In het geval van de Haïti studie kan worden gesteld dat dit rapport geschikt is om bij aan te haken wanneer in 2015 de meta evaluatie van de Nederlandse noodhulp moet worden gedaan.

Met betrekking tot de evaluatie van Sudan kan worden geconstateerd dat voor de eerste doelstellingen (betreffende impact van de interventies en aanbevelingen voor de periode na 2010) de relatie tussen analyse/conclusies en de doelstellingen van de evaluatie adequaat uiteengezet is. Wat betreft het laatste doel (analyse van de toepasbaarheid van de CPPB-benadering) geldt dat deze benadering voldoet als analytisch kader dat nadere invulling behoeft in concrete toepassingen. Wat betreft de zeer gemengde resultaten betreffende de effectiviteit en efficiency van de verschillende mechanismes van hulp geldt dat de bevindingen goed geadstrueerd lijken te zijn.

Het rapport over capaciteitsontwikkeling – een tot dusverre niet of nauwelijks systematisch bestudeerd deel van ontwikkelingssamenwerking – heeft op een aantal punten belangrijke conclusies en aanbevelingen geformuleerd. Veel van de conclusies en lessen illustreren dat een dergelijke studie een nuttig leereffect kan hebben. Het syntheserapport draagt bij aan de doordenking van de betekenis van 'endogeniteit' bij capaciteitsontwikkeling en biedt ook nieuwe kennis over de bijdragen van Nederlandse partners aan die processen bij Zuidelijke organisaties en systemen. Het is echter niet altijd duidelijk waar de conclusies voortvloeien uit de data en waar IOB reflecteert. De referenties naar 'Dutch expertise' en 'DDP support for innovation' in de conclusies en lessen zijn gezien het referentiekader en de scope van het onderzoek pre-matuur: er is immers geen vergelijkend onderzoek gedaan naar de performance van andere ontwikkelingspartners. Dat DDP-support toch vaak op improductieve wijze het proces van capaciteitsontwikkeling bepaalde kanten uittrekt die de duurzaamheid ervan kan schaden, is een bevinding die blijkbaar nog steeds relevant is – ondanks dat ook IOB al vaak op dit soort risico's heeft gewezen.

Met de nogal breed opgezette aanpak in het tweede hoofdstuk van het synthese rapport over capaciteitsontwikkeling (de setting van capaciteitsontwikkeling en donorstun aan de hand van documentatie) kwamen

zaken aan bod die niet direct voortvloeien uit de vraagstelling die aan het rapport ten gronde lag, zoals de kwaliteit van het Nederlandse beleid inzake capaciteitsontwikkeling en de effectiviteit van bepaalde beleidsinstrumenten inzake ondersteuning van medefinancieringsorganisaties. De kwaliteit van het Nederlandse beleid viel niet binnen de ToR maar er worden wel uitspraken over gedaan, waarvan sommige in de beleidsreactie zijn overgenomen en daarmee beleid worden. Ook bij de evaluatie van de twinning faciliteit Suriname-Nederland werd getracht een aanzet tot een antwoord te geven op een vraag die niet in de ToR (beleidskader) stond (zie boven). De juistheid van het op deze manier door de mazen van een vraagstelling heen kruipen en andere (vaak op zich best legitieme) vragen aan de orde stellen, roept vragen op. In methodologische termen kan men zich bijvoorbeeld afvragen of het onderzoek wel voldoende 'evidence' heeft kunnen opleveren voor het beantwoorden van die verderstreckende vragen. Wel kan een rapport in een metadiscussie op het eind de kaders wat breder trekken en verderstreckende vragen aan de orde stellen, ook al kunnen die strikt genomen niet beantwoord worden.

2.6 Handvatten voor stakeholders

Het gaat hier om de relatie tussen bruikbaarheid en gebruik, anders gezegd om de bruikbaarheid van evaluaties vanuit het perspectief van de verschillende categorieën van belanghebbenden. In het kader van de leerfunctie van evaluaties moeten de bevindingen kunnen leiden tot het zoeken naar en het introduceren van opties die de doelmatigheid en de doeltreffendheid van de onderzochte activiteiten verbeteren. Daarmee is het van belang of er in het rapport aanzetten worden gegeven tot een dergelijk *gebruik*. Dergelijke lessen (of aanzetten daartoe) kunnen in het rapport zelf zijn opgenomen of kunnen worden getrokken door het rapport in workshops met belanghebbenden te bediscussiëren.

In het geval van de NIMD evaluatie (no. 331) wordt gesteld dat door de veelheid van sets van conclusies en lessen, de lezer moet zoeken naar aanzetten wat de bruikbaarheid niet ten goede komt. In het geval van capaciteitsontwikkeling (no. 336) worden stakeholders, waaronder ook zuidelijke organisaties, expliciet geadresseerd. Tijdens het evaluatieproces werd een workshop georganiseerd met de vertegenwoordigers van de Nederlandse organisaties, echter zonder directe inbreng van de zuidelijke partners. In de evaluatierapportage over MFS-I (no. 362) wordt aan dit punt geen aandacht geschonken. De Zuid-Sudan studie gaf aanbevelingen aan een reeks van instanties: donoren, de bredere hulp-gemeenschap, de ontvangende overheden. Tijdens een workshop met stakeholders werden de conclusies en aanbevelingen besproken. Ook de Congo-evaluatie geeft onder een drietal hoofdthema's telkens specifieke aanbevelingen voor donoren, agentschappen en NGOs.

3. Gebruik bestudeerde IOB evaluaties

Het gebruik van IOB evaluaties is een onderwerp waar het PVA nog onvoldoende inzicht in heeft. Gebruik in de beleid(sbepaling, resp. uitvoering) hangt ook samen met de kwestie van de institutionele verankering, waar dit advies van het Panel niet over gaat. Er is echter wel een aantal factoren, dat het gebruik van de evaluaties beïnvloedt, respectievelijk zichtbaar maakt.

3.1 Zichtbaarheid van de rapporten

Als het gaat om zichtbaarheid van IOB-rapporten is vooral gelet op de zichtbaarheid op het World Wide Web van de rapporten en eventuele reacties daarop in de vorm van recensies, discussies, follow-up-activiteiten als workshops, e.d. De IOB-evaluaties zijn te vinden via de zoekmachine google. Op de website van het ministerie van Buitenlandse Zaken zijn de rapporten (met eventueel de beleidsreactie en evaluatienieuwsbrief als het gaat om IOB-publicaties) ook te vinden, maar na enig zoekwerk. Het is overigens in de lijst van afgeronde evaluaties op het eerste gezicht niet voor iedereen helder of de publicatie een IOB-rapport is of een decentraal evaluatie-rapport. Naast de online-wereld worden IOB-rapporten fysiek tentoongespreid. Indien opportuun worden ze gepresenteerd tijdens lunchlezingen en seminars op het ministerie of bij *events* buiten het ministerie, zoals de

Afrikadag georganiseerd door de Evert Vermeer Stichting. Van deze laatste vorm van verspreiding heeft het PvA geen volledig beeld.

3.2 Toegankelijkheid

Ten opzichte van het Eerste Advies is er een initiatief genomen om de toegankelijkheid van IOB-rapporten te vergroten. In 2011 is de IOB-evaluatienieuwsbrief geïntroduceerd. Met weinig woorden en in 'gewone' taal probeert IOB op deze wijze essentie van een rapport helder te communiceren naar het bredere publiek. Deze *two-pager* wordt standaard samen met het hoofdrapport en de beleidsreactie naar de Tweede Kamer en andere afnemers verstuurd. Het PvA constateert het panel dat deze evaluatienieuwsbrieven een heldere en smaakmakende snelle toegang verlenen tot het rapport. De voorwaarde om de boodschap toegankelijker te maken lijkt daarmee te zijn gerealiseerd. Of dit uiteindelijk heeft geleid tot meer mensen die kennis nemen van de inhoud van IOB-evaluaties, is niet onderzocht.

Wat de toegankelijkheid van de rapporten zelf betreft, concludeert het PvA dat de structuur van de rapporten doorgaans helder is en het feitenmateriaal op inzichtelijke wijze wordt gepresenteerd. Ook het gebruik van Annexen draagt bij aan een toegankelijke ordening van het materiaal. Daarbij is in de gezamenlijke evaluaties gebruik gemaakt van presentatie van het materiaal op Cd-Rom. De teksten zijn overigens niet altijd even lezersvriendelijk geformuleerd; meer aandacht daaraan zou de leesbaarheid ten goede kunnen komen en zo de toegankelijkheid vergroten.

3.3 Feitelijk gebruik

Tijdens de uitvoering van evaluaties worden ToR's, deelrapporten en eindrapport (inclusief voorlopige bevindingen) teruggekoppeld naar stakeholders en externe professionals voor een hoor-wederhoor moment. Na vaststelling van het evaluatierapport door directeur IOB, wordt dit rapport en evaluatienieuwsbrief voorzien van een beleidsreactie van de betrokken bewindspersoon aan het parlement gezonden. De reactie geeft aan welke maatregelen de bewindspersoon neemt op basis van de conclusies en aanbevelingen van IOB. Leden van de vaste Kamercommissie Buitenlandse Zaken kunnen indien gewenst schriftelijk vragen stellen of een algemeen overleg vragen over de beleidsreactie en de evaluatie. Directie Financieel Economische Zaken (FEZ) is verantwoordelijk voor het bijhouden van een overzicht van de naleving van de maatregelen door directies.

Feitelijk gebruik is in dit Tweede Advies voornamelijk beperkt tot opmerkingen die OS-woordvoerders in de Tweede Kamer hierover maakten in hun gesprekken met leden van het Panel. Uit de gesprekken bleek zowel dat de Kamerleden de rapporten als nuttig ervaren bij hun werk als dat ze deze op een voor hen opportune wijze gebruiken. Hierbij wordt geen onderscheid gemaakt tussen deelstudies of synthese studies. Beide kunnen voor de voorbereiding van een parlementair debat van belang zijn. Het is niet aan IOB om een selectie te maken, door bijvoorbeeld niet de onderliggende deelstudies maar slechts de beleidsdoorlichting op een bepaald terrein toe te sturen. De verschillende Kamerleden gebruiken de rapporten elk op hun eigen manier, waarbij steeds een politieke afweging wordt gemaakt. Het algemene gevoel is dat het de taak van IOB is om objectieve en goede rapporten maken. De Kamer leden kunnen er mee doen wat voor hun politiek opportuun is. De regel is dat beleidsdoorlichtingen altijd naar de Kamer worden gestuurd, terwijl effectenonderzoeken en andere evaluaties worden gestuurd als daar evidente argumenten voor zijn, bijvoorbeeld als de Kamer daarom vraagt. Uit de gesprekken met enkele OS-woordvoerders in de Kamer is echter niet gebleken dat hun afweging van politieke relevantie en opportuunheid van IOB-rapporten deze tweedeling volgt.

De beleidsreactie van de betrokken bewindspersoon en de parlementaire discussie over de beleidsreactie en het rapport zijn vastgelegd, maar het PvA ziet het niet als zijn taak daarvan in dit advies een overzicht te geven. Uit de informatie die IOB hierover aan het panel verstrekke, blijkt dat in sommige gevallen het rapport en de beleidsreactie veel aanleiding tot discussie gaven in de Tweede Kamer, zoals het rapport over de Twinning faciliteit Suriname-Nederland (no. 361), terwijl dit voor andere rapporten veel minder het geval was. Sommige

van de door het panel bestudeerde rapporten zijn meerdere keren onderwerp van discussie geweest, zoals het NIMD rapport (no. 331) en de beleidsdoorlichting Basic education (no. 353). Hieruit blijkt dat de rapporten een nuttige functie vervullen in het politieke debat, of het nu gaat om de bilaterale relatie met Suriname (voortzetting van de twinning faciliteit in het licht van de politieke ontwikkelingen in Suriname) of om een OS sector waarvan de Nederlandse bijdrage ten principale ter discussie staat (basis educatie).

De doorwerking van het rapport in het beleid van het ministerie van Buitenlandse Zaken en de opvolging van toegezegde maatregelen door de betrokken bewindspersoon vallen onder de verantwoordelijkheid van FEZ. Het punt van de inhoudelijke doorwerking van evaluaties in het beleid maakt geen deel uit van het mandaat van het PvA en is daarom geen onderwerp van onderzoek geweest. Ook de doorwerking van het rapport onder stakeholders (partnerlanden, NGOs etc.) is niet systematisch bestudeerd. Wel is getracht na te gaan (via analyse van internet-hits) of er sprake was van zulke manifestaties van gebruik, bijvoorbeeld in de vorm van besprekingen en discussies, traceerbare verwijzingen, et cetera. Van zulke manifestaties was slechts sporadisch sprake.

4. De evaluatieprogrammering

4.1 Waarom dit thema?

Het ministerie van Buitenlandse Zaken /IOB heeft een evaluatieprogrammering opgesteld voor alle beleids-terreinen. Deze programmering bestrijkt een periode van zes jaar en wordt jaarlijks bijgesteld en aan de Tweede Kamer aangeboden samen met de begroting. Onder programmering wordt verstaan het selecteren en categoriseren van onderwerpen voor evaluaties die in uitvoering genomen worden en waar middelen voor beschikbaar zijn, en waarbij als uitgangspunt geldt dat elk beleidsterrein op het niveau van de operationele doelstellingen (ODs) (uit de begroting) periodiek onderwerp is van studie. Studies kunnen evaluaties zijn maar ook beleidsdoorlichtingen of overzichtsstudies. Dit conform de Regeling Prestatie indicatoren en evaluaties 2006 (RPE). Naast de RPE 2006 kunnen er ook andere overwegingen zijn om evaluaties in de programmering op te nemen. Het PvA heeft naar de uitgangspunten van de IOB programmering, de wijze waarop deze tot stand komt en de uitkomsten van het proces gekeken omdat:

- De programmering d.m.v. het agenderen van bruikbare evaluaties het gebruik van evaluatierapporten moet bevorderen zodat de leerfunctie van evalueren kan worden gerealiseerd;
- De programmering een instrument is om inzicht te geven in en verantwoording af te leggen over de doelmatigheid van de inzet van middelen voor het bereiken van de doelstellingen van ontwikkelings-samenwerking, de verantwoordingsfunctie van evalueren (accountability).
- Ons signalen bereikten uit OECD kringen dat zich bijvoorbeeld in Engeland ingrijpende veranderingen voordoen, resp. hebben voorgedaan, ten aanzien van de wijze van programmering én ten aanzien van de positionering van evaluatiediensten van de overheid op OS gebied. Daar kennis van nemen en vergelijkingen proberen te maken met de Nederlandse situatie achtte het Panel van Advies belangrijk.

4.2 Hoe heeft het panel dit vraagstuk aangevat?

Door (1) beperkt (vier landen) vergelijkend onderzoek door een externe deskundige te laten doen (Helmsing, 2012), (2) een viertal Kamerleden (OS woordvoerders) te bevragen over de invloed van het Nederlandse parlement op de programmering, (3) enkele IOB medewerkers te bevragen over het programmeringsproces en eventuele aantekeningen die ze daarbij hebben, alsook (4) via een *quick scan* na te gaan wat in de literatuur over het onderwerp te vinden is. Het onderstaande is gebaseerd op deze vier bronnen.

4.3 Bevindingen

4.3.1 Relevante literatuur

Onze zoekacties naar literatuur leverden zeer weinig op. Wat er is, is verouderd (vgl. van Hoesel, 1985), vooral wetenschapstheoretisch van aard (Lakatos, 1978) of is uitsluitend gewijd aan aspecten van coördinatie van

onderzoeksactiviteiten ('de planning' en welke methoden en aanpakken gebruikt worden). Het overgrote deel van de verwijzingen gaan over 'program evaluations' in plaats van over 'programming of evaluations'. Een kleiner deel gaat in op het evalueren van hoe beleidsprogramma's tot stand komen, maar ook dat is niet wat hier bedoeld is. Iets meer aandacht is er voor nationale (top) programma's van wetenschappelijk onderzoek, als ook wat de EU op dit gebied doet¹⁰. Echter, dat 'niveau' onderzoeksprogrammering omvat zo goed als geen specifieke evaluaties en is van een ander 'level playing field' dan waar het in ons geval om gaat.

De in het kader van de analyse van enkele relevante buitenlanden door Helmsing (2012) ondervraagde deskundigen leverden een interessant beeld op. Meermalen werd hij geconfronteerd met 'weary eyes' als hij vragen stelde over de aard, achtergronden en uitgangspunten van de programmering van evaluaties. Ook vielen hem commentaren ten deel zoals "that is a very good question" (Deense geïnterviewde), dan wel beschouwingen zoals dat programmering "is more an art than a science" (Zweedse geïnterviewde), of dat programmering een "een heuristisch proces is met en van verstandige mensen" (Nederlandse geïnterviewde). Een Britse geïnterviewde zei dat het programmeren van evaluaties in de OS wereld een "unashamedly unscientific" proces is.

4.3.2 Het vergelijkend onderzoek

Het vergelijkende onderzoek van Helmsing richtte zich vervolgens op de programmering in drie landen: het VK, Zweden en Denemarken. Onderstaande tabel geeft aan om welke organisaties het gaat en waar contact mee geweest is. Er is daarbij onderscheid te maken tussen drie niveaus waarop diensten opereren, namelijk het externe evaluatie agentschap (ICAI en SADEV), de centrale evaluatie unit binnen een departement (IOB, SIDA, evaluatie unit van Denemarken, DFID) en decentrale evaluatie-activiteiten. Aangetekend moet worden dat de gegevensverzameling medio 2012 werd afgerond. Recentere ontwikkelingen zoals de oprichting van het nieuwe Duitse Deutschen Evaluierungsinstituts der Entwicklungszusammenarbeit (DEval)¹¹ en de 'ontmanteling' van SADEV zijn derhalve niet meegenomen. Dat zelfde geldt voor een direct voor de programmering van evaluaties relevante activiteit binnen DGIS, Quality at Entry (Q@E), die de evalueerbaarheid van beleid poogt te verbeteren.

Tabel 1. Overzicht van niveaus waarop evaluaties van OS gericht zijn in vier landen (Helmsing, 2012)

	VK	Denemarken	Zweden	NL
Onafhankelijke evaluatie unit	ICAI	---	SADEV	---
Centrale Evaluatie-unit bij een departement	(*)	Evaluatie departement Ministerie van Buitenlandse Zaken	Evaluation & Monitoring Unit SIDA	IOB / ged. Ministerie van Buitenlandse Zaken
Gedecentraliseerde evaluatie(units) van ODA organisaties en 'field offices'	'embedded' in DFID programma teams en landenkantoren/organisaties	----	Beleidsteams & ambassades	Directies, beleidsafdelingen, ambassades

*DFID heeft een evaluatie unit, maar voert geen evaluaties meer uit.

Binnen IOB wordt er aan de hand van de begrotingsartikelen een programmering opgesteld van onderwerpen die over een bepaalde periode voldoende dekking moeten geven van deze artikelen. Verder zet IOB ook thema's op de programmering die dwarsdoorsnijdend zijn (zoals coherentie, kanaalvergelijking, capaciteits-

¹⁰ http://ec.europa.eu/research/era/areas/programming/joint_programming_en.htm;

http://ec.europa.eu/research/press/2008/pdf/com_2008_468_en.pdf

¹¹ Dit instituut heeft de volgende taken. " Das DEval prüft, welche Maßnahmen etwas bewirkt haben und welche nicht. Es macht Vorschläge, was man zukünftig bei ähnlichen Projekten verbessern kann. Und es gibt Wissen und Erfahrungen an andere weiter.

ontwikkeling) of brede, 'ontschotte' thema's (zoals de Afrika evaluatie, en sectorstudies). Tot slot kijkt het ook naar onderwerpen in relatie tot het maatschappelijk debat.

In het *Verenigd Koninkrijk* is de focus van ICAI gericht op actuele programma's en het management ervan. Het gaat vooral om evaluaties die 'real time' plaats vinden in plaats van ex-post. De focus van die studies is 'issue-gericht': is OS kosteneffectief? Is de steun daadwerkelijk gericht op de behoeften van arme mensen? Voor 2012 heeft ICAI bijvoorbeeld evaluaties geprogrammeerd van DFID's onderwijsprogramma's in drie Oost-Afrikaanse landen en van het management van de zog. 'Cross-Departmental Conflict Pools'. ICAI evalueert ook direct multilaterale agentschappen in plaats van zich te beperken tot het reviewen van de multilaterale evaluatie-functie van deze organisaties. Een evaluatie van een regionale ontwikkelingsbank is ook geprogrammeerd voor 2012 (Flex-component). In deze evaluaties is de centrale vraag: hoe draagt DFID er zorg voor dat multilaterale organisaties zich zodanig engageren dat het realiseren van de UK OS doelen gewaarborgd is?

In *Zweden* is de SADEV, conform haar mandaat, in haar programmering gericht op de landelijke nationale samenwerkingsstrategieën en op multilaterale activiteiten. Voor de komende jaren staat de evaluatie van de ontwikkelingssamenwerking in conflictsituaties op de agenda. De Central Evaluation Unit van SIDA heeft voor 2012 een aantal lijnen uitgezet voor (andere) strategische beleidsevaluaties, te weten:

- 'Actors in development' (particuliere sector, maatschappelijk middenveld en parlementen) en
- Institutionele ontwikkeling, 'governance' en economische groei (het beheer van gronden, milieu-support en *governance, empowerment* van de jeugd op economisch gebied, elektrificatie voor mensen in armoede en capaciteitsopbouw).

In *Denemarken* heeft de evaluatieafdeling van het betreffende ministerie een strategische beleidsevaluatie van de DANIDA Civil Society-strategie geprogrammeerd en als onderdeel van dit plan thematische studies (naar seksuele en reproductieve gezondheid en rechten; HIV / AIDS) voorzien. Een ander object van een thematische evaluatie heeft betrekking op de Deense steun aan 'fragile states'. Modaliteiten van ontwikkelingshulp zijn het derde thema waar thematische evaluaties zich op richten, in het bijzonder beheer van de overheidsfinanciën en begrotingssteun.

4.3.3 Kernthema's

Uit het vergelijkend onderzoek en uit gesprekken die in Nederland gevoerd zijn, blijken de volgende factoren samen te hangen met het proces en de inhoud van de programmering:

- a) De localisatie of te wel positionering van de evaluatiefunctie en belangrijke veranderingen daarin;
- b) Verantwoordelijkheid voor de besluitvorming t.a.v. inhoud van de programmering;
- c) Het gewenste evenwicht tussen de leerfunctie en de accountability functie [inclusief de 'coverage' die vanuit de FEZ directie/ RPE nodig is van evaluaties] en de mogelijke invloed van de politiek hierop;
- d) De mate waarin externe stakeholders en het 'maatschappelijk debat' bij de inhoud van de programmering worden betrokken;
- e) Overwegingen van doelmatigheid van de evaluatiepraktijk.

Deze thema's lichten we nu toe.

Ad a: de localisatie van de evaluatiefunctie en gebruik van evaluaties

Evaluaties vinden plaats op diverse 'niveaus' binnen publieke sector organisaties. Gaat het om het operationele niveau, dan beogen evaluaties primair bij te dragen aan instrumentele beleidsontwikkeling en –uitvoering. Gaat het om de meer centrale niveaus, dan staan evaluaties vooral in dienst van wat Helmsing (2012) "policy control" noemt, "which broadly aligns with a conceptual use for policies but it also serves internal accountability as means of control". Wanneer het gaat om externe en/of geheel onafhankelijk uitgevoerde evaluaties, dan, zo stelt Helmsing (2012) beogen evaluaties vooral bij te dragen aan het overtuigen van de centrale overheid en de samenleving van "the legitimacy of the evaluand, in our case, ODA. In recent years

public accountability has become an important use of ODA related evaluations: convincing the public in order to maintain broad based political support”.

Een tweede element dat Helmsing op het spoor kwam is dat van de ‘vertical alignment between different levels of evaluations’. Voor Nederland wordt in dat verband gewezen op het fenomeen van de dakpansgewijze programmering bij het Ministerie van Buitenlandse Zaken en IOB. De kern ervan is dat de evaluaties op het operationele niveau inzichten en bevindingen produceren waarop, idealiter, voortgebouwd kan worden door de ‘hogere’ in de organisaties plaats vindende evaluatieve activiteiten die als ‘policy control’ werden beschreven. Helmsing: “Based on evidence from various policy areas of ODA the overall use and results of official development assistance can be evaluated for the purpose of legitimizing ODA. The assumption is that a careful roof tile construction when being involved in programming evaluations will enhance the latter’s effectiveness and efficiency”.

Helmsing ging na of deze dakpansgewijze programmering ook in de drie andere landen voorkwam. Dat bleek maar gedeeltelijk het geval te zijn. “In the UK the creation of the external evaluation agency triggered the removal of the central evaluation unit as an organization responsible for policy control work. At the same time the decentralized ‘embedded evaluation’ function has not limited itself to operational uses (reviews) but has also been involved in policy and thematic evaluations. However, in Sweden the creation of SADEV did not have similar effects”. De Danish Evaluation Department (EVAL) van het Ministerie van Buitenlandse Zaken is verantwoordelijk voor planning, management en ‘quality control of evaluations’. Evaluaties die decentraal aangevat worden maar er overigens niet of nauwelijks zijn, richten zich meer op operationele topics. “That is to say it is undertaken as part of programme and project cycles (mid-term reviews and end of term evaluations to support management decisions)”. Ambassades verrichten geen evaluaties noch laten ze die doen, zij het dat “the Evaluation department invites embassies to propose evaluations”.

In het VK, Zweden en Nederland is de figuur van decentrale evaluaties er uitdrukkelijk wel maar er zijn wel belangrijke verschillen tussen deze landen. In Nederland laten (thematische) directies evaluaties uitvoeren, soms door IOB, soms door derden, waarbij overigens aangetekend zij dat niet alle studies ook onderwerp van de programmering zijn. Evaluaties van en door Ambassades “are not systematically covered within the evaluation system. In Sweden decentralised evaluations are labelled ‘reviews’ and can be found on the Sida’s evaluation website. There is some central use made of these. In the UK the creation of the ICAI was an important external development, which triggered within DFID an equally important internal change. The high level external scrutiny caused by the establishment of ICAI together with the effects of the work of IACDI stimulated DFID senior management to pay more attention to evidence based policies and programmes.”

Ten aanzien van de situatie in het VK benadrukt Helmsing (2012) ook het volgende. “A second important difference is that DFID has taken the programming of evaluation out of the central evaluation unit and placed it in the decentralized units through embedding of evaluation in policy and program design and implementation. Thus, effectively there is no central organizational level programming anymore. The central evaluation unit is supporting decentralized units to successfully incorporate evaluation in policy and program design by setting quality standards and by providing evaluation capacity. Ownership of evaluation has shifted from the central evaluation unit to the decentralized policy and program teams. The types of evaluation that are programmed at decentralized levels are not restricted to operational uses (reviews) but include also policy, thematic and impact evaluation alongside process evaluations and other studies.”

Helmsing (2012) heeft ook, tegen de achtergrond van het concept van ‘disjointed incrementalism’¹², naar voren gebracht dat de landen verschillen in de mate waarin er ‘organisch’ wordt omgegaan met de

12 Lindblom benadrukte dat “it is not irrational for an administrator to defend a policy as good without being able to specify what it is good for” en merkte ook op dat “agreement on policy thus becomes the only practicable test of the

programmering. Bestaat er in het Verenigd Koninkrijk (DFID) het 'blue book', "which describes organizational routines and policy, program and project procedures (OECD/DAC, 2011) and where policy control would seem more hierarchical and the organization to be more rule driven", in Nederland is meer sprake van een 'organisch' proces van totstandkoming van de programmering. Binnen het Ministerie zijn vele tientallen, zo niet meer budgethouders die de ontwikkeling van beleid, beheer en controle voor hun rekening nemen (alsook evaluaties). Tal van personen en teams, bestaande uit medewerkers van de diverse eenheden, zijn in een (gedurig) proces van onderhandelen en afstemmen betrokken. "The policy formulation is more 'organic' that is to say; policy notes can be prepared in different corners of the organization and are not controlled by a central policy formulation unit"¹³.

Zonder meer mede bepalend voor wat aan programmeringsactiviteiten verricht wordt binnen de OS wereld is dat in twee van de vier landen recent externe Evaluatie-'agencies' zijn opgericht (VK en Zweden). Tegelijkertijd dient beseft te worden dat de rationale om daar toe te besluiten fundamenteel verschillend is. "In Sweden it was motivated by an administrative reform where the aid sector was brought in line with other sectors of the Swedish public administration of semi-autonomous implementation agencies and external evaluations agencies to evaluate their performance, while in the UK public accountability of aid was main driver in order to maintain public support for rising aid budget". Wat betreft het VK, spreekt Helmsing over een 'regime change'. Kenmerkend daarvoor is dat evaluaties die primair gericht zijn op 'accountability' buiten DFID geplaatst zijn bij ICAI en dat daarmee ook een verandering van de aard en onderwerpen van de evaluaties heeft plaatsgevonden (meer 'real time' en meer gericht op value-for-money). "At the same time, this policy regime change created the room for an important organizational and management change within DFID. By embedding evaluation in policy and program teams, the programming of evaluation was also decentralized. The human resources and technical capacity of central evaluation unit, which came available, was redeployed (and expanded) to support this".

De afgelopen jaren hebben zich op het terrein van IOB ook de nodige veranderingen voorgedaan in het beleidskader van de evaluatieplanning:

- versterkte interministeriële beleidsvoering; vooral op terreinen van internationaal beleid inzake landbouw en bedrijfsleven (voormalig EL&I), milieu (I&M), migratie (BZK);
- ontwikkelingen binnen de RPE (evaluaties worden in toenemende mate opgezet als een traject van integrale beleidsevaluatie op het niveau van OD's; een nieuwe RPE is 'in the making');
- analyses van de Algemene Rekenkamer en WRR over de rol van effectonderzoek en het gebruik van evaluaties;
- afkondiging van het Integraal Afwegingskader (IAK) gericht op de inpassing van de resultaten van het evaluatieonderzoek in de beleidscyclus van het ministerie;
- methodische ontwikkelingen op het vakgebied (bv. impactstudies, real-time evaluaties, randomized controlled studies/ trials, etc.).

De verschillende geïnterviewden geven aan, dat deze veranderingen invloed hebben op de programmering.

Ad b: Verantwoordelijkheid voor de besluitvorming t.a.v. inhoud programmering

Er bestaan verschillen tussen landen in de wijze waarop het proces van de evaluatieprogrammering wordt vormgegeven en uitgevoerd. Denemarken heeft sinds een lange periode (de jaren zestig) een neo-corporatistische traditie waarin het bestuur van DANIDA werd betrokken in besluitvorming van beleid ten aanzien van de hulp en in de evaluatieprogrammering (sinds de jaren negentig). In het VK heeft, zoals gezegd, de recente verandering in het beleidsregime geleid tot een radicale wijziging in de besluitvorming en het proces van programmering.

policy's correctness" (Lindblom in Gawthrop, 1970: 211). Een en ander zou zich niet wel verdragen met een meer rationele benadering die centraal staat bij evaluatie(onderzoek), waar het juist wel gaat om 'why what is good [or bad] for'.

13 Over Zweden en Denemarken heeft Helmsing geen informatie verzameld.

Binnen de beperkingen van de studie was het niet mogelijk om de verschillen in de wijze van programmeren te attribueren aan de specifieke rollen/rolverdeling van bepaalde stakeholders (denk aan interne bureaucratie, politieke en sociale actoren zoals Parlement of NGOs). Ambtenaren en academici in Denemarken erkennen de rol van de leiding van DANIDA in de programmering. Deze erkenning heeft medewerkers van DANIDA de mogelijkheid gegeven om bepaalde thema's voor te stellen die in de evaluatie(s) aandacht krijgen. De leiding van DANIDA pleitte bijvoorbeeld sterk voor het doen van gezamenlijke evaluaties. Uit de interviews bleek dat het advies van de leiding van DANIDA de legitimiteit vergrootte van de evaluatieprogrammering en dat hierdoor de programmering vervolgens redelijk snel zijn weg vervolgde door de bureaucratie en langs het Parlement.

Directeur IOB heeft de eindverantwoordelijkheid voor de evaluatieprogrammering op het terrein van ontwikkelingssamenwerking; op de overige beleidsterreinen van het buitenlands beleid adviseert IOB over de programmering en ligt de eindverantwoordelijkheid bij de betreffende directeuren-generaal. De vaststelling van de verdeling van verantwoordelijkheden heeft zijn beslag gekregen in de afgelopen jaren waarin de rol van de Inspectie opnieuw onderwerp van politiek en ambtelijk debat is geweest. Daarbij kwamen veelal dezelfde opvattingen aan de orde als tijdens de jaren 80 van de vorige eeuw, zoals de onafhankelijkheid van evaluatie, de openbaarheid van rapporten, de verhouding tussen externe verantwoording en intern leren. Dit heeft ertoe geleid dat de (toenmalige) minister voor ontwikkelingssamenwerking in 2008 de onafhankelijkheid van de Inspectie en de openbaarheid van haar rapporten tegenover de Kamer heeft herbevestigd. Wel werd de programmeringsbevoegdheid van IOB beperkt tot ontwikkelingssamenwerking. De evaluatieprogrammering van het ministerie van Buitenlandse Zaken wordt opgenomen in de Memorie van Toelichting, is openbaar en kan worden gevonden op de website van het Ministerie van Buitenlandse Zaken/rijksoverheid.

Door (geïnterviewde) Kamerleden wordt bij de vraag hoe zij naar de programmering van onderzoek door IOB kijken gerefereerd aan COS/AIV, als het om agenderen van specifieke onderwerpen gaat. Het argument is dat de AIV onafhankelijk is (i.e. buiten het apparaat van Buitenlandse Zaken staat). Weliswaar doet de AIV verkennende beleidsstudies en geen evaluaties, maar in de adviesfunctie zit een behoorlijk deel 'beleids-evaluatie'. Volgens een van de geïnterviewde parlementsleden gaan Kamerleden daarom eerder naar AIV (COS) toe als ze onderwerpen op de agenda willen hebben dan naar IOB.

Wat verder is waar te nemen is dat er sprake is van een toenemende verschuiving van de evaluatiepraktijk van decentraal naar IOB. Deze tendens heeft consequenties voor de stafcapaciteit van IOB. Door te benadrukken dat baseline gegevens van belang zijn om (effect)evaluaties 'überhaupt' op professionele en doelmatige wijze te kunnen uitvoeren, heeft IOB beleidsmedewerkers geprobeerd ervan te overtuigen dat baselines nodig zijn bij aanvang van een programma opdat men het programma later kan monitoren en evalueren.

Er is sprake van een 'voortrollende' programmering met jaarlijkse bijstelling. IOB en FEZ coördineren deze programmering en plegen overleg met directies over de thema's in het betreffende werkkerrein die zich lenen voor een evaluatie of beleidsdoorlichting. IOB houdt zich bezig met beleidsevaluaties, terwijl directies en ambassades meer geïnteresseerd zijn in de effecten van de interventies en effectiviteit van de activiteiten (doelmatigheid van het proces), niet zo zeer in de (langere termijn) effecten en impact van beleid.

Ad c: Het gewenste evenwicht tussen de leerfunctie en de accountability functie

Er lijkt sprake te zijn van een zekere verschuiving richting de accountability functie mede onder invloed van de politieke druk om ODA gelden helder te verantwoorden. In ieder geval heeft dit duidelijk een rol gespeeld in de veranderingen van de evaluatiepraktijk en programmering in het VK. In een van de interviews met IOB staf werd opgemerkt dat de relatie tussen bruikbaarheid en programmering lastig is, als gevolg van de kloof tussen de beleidscyclus en de vaak gebrekkige operationalisering van beleid. Als dit een juiste observatie is, zou dat ten koste gaan van de leerfunctie van evaluaties.

Ad d: De mate waarin externe stakeholders en het ‘maatschappelijk debat’ bij de inhoud van de programmering worden betrokken

De landenvergelijking levert op dat de mate van *betrokkenheid van externe stakeholders* bij de formulering van de onderzoeksprogramma's behoorlijk verschilt. Bij het Britse (ICAI) driejaarlijkse werkprogramma spelen externe stakeholders een belangrijke rol maar Helmsing (2012) merkt er wel bij op dat "a cynical view would be that after having widely consulted, the ICAI could justify any pre-determined selection by invoking such consultations. A more constructive view would be that since public accountability is the key factor, it makes sense for the ICAI to cast its harvesting of information as wide as possible and not limit itself to information provided by the evaluand (DFID)".

Overigens moet hij ook vaststellen dat ook in situaties waar externe stakeholders geen directe rol hebben bij de programmering van evaluatieonderzoeken, "directors of evaluation units acknowledge the importance of scanning the organization's environment for signals from key stakeholders such as politicians, senior administrators, influential academics and practitioners, etc., even if the director might have the formal mandate to decide independently on the programming".

Dit blijkt ook in de praktijk van IOB. Een belangrijke stap in het proces om te komen tot een degelijke evaluatie programmering van het Nederlands buitenlands beleid zijn de consultaties die IOB houdt met directeuren en beleidsmedewerkers van het departement. Er wordt verder gekeken naar de RPE criteria, de vraag vanuit de politieke en ambtelijke lijn, en de maatschappelijke relevantie. Een van de primaire stakeholders van IOB is de Tweede Kamer. Actieve invloed van de Kamer op de programmering wordt zelden uitgeoefend. De Afrika evaluatie was een voorbeeld waar bij dat wel gebeurde. Volgens een geïnterviewde hoeft de programmering niet per definitie aan te sluiten bij het maatschappelijke en politieke debat. Dat is niet aan het IOB. Die moet inhoudelijk goede rapporten maken en verantwoording afleggen over het gevoerde beleid en de bestede belastinggelden. NGO's hebben geen directe invloed op de programmering.

Ad e: Overwegingen van doelmatigheid van de evaluatiepraktijk

IOB volgt de richtlijnen van de RPE en probeert zoveel mogelijk beleidsartikelen af te dekken. Overigens zijn deze artikelen zo breed geformuleerd dat er veel onder valt. Echter daarnaast is IOB van mening dat het volgen van de beleidsartikelen en het streven naar voldoende 'coverage' in overleg met FEZ en volgens de RPE, door de verantwoordingsverplichtingen niet als een keurslijf moet worden opgevat. De RPE schrijft het belang voor van beleidsdoorlichtingen. IOB voert deze ook uit, maar daarnaast programmeert het ook impact evaluaties, 'systematic reviews' en andere studies. IOB voert zelfstandig evaluaties uit en adviseert over de evaluaties van beleidsdirecties op alle beleidsterreinen die vallen binnen de homogene groep buitenlanduitgaven (HGIS). Het evaluatiewerk van IOB heeft zich de laatste jaren ontwikkeld van projectevaluaties naar brede onderzoeken op sector-, thema- en landen/regioniveau. Een belangrijk deel van de beleidsdoorlichtingen op het niveau van operationele doelstellingen is thans gebaseerd op ex-post effectenonderzoek met toepassing van statistische methoden voor impactevaluatie.

ICAI is hoofdzakelijk actief met Value for Money (VFM) reviews en studies. ICAI richt zich niet op impact studies op grond van de overweging dat dit soort evaluaties een ex post karakter hebben. In plaats daarvan wordt prioriteit gegeven aan evaluaties die gericht zijn op het verbeteren van de basis voor succes tijdens de beleidsimplementatie (eerder korthedshalve als 'real time' –evaluaties aangeduid). Daarbij spelen VFM studies een belangrijke rol. "An important new and different development in the UK is the embedded evaluation in DFID whereby evaluations are programmed at decentralized policy and program formulation stage by respective policy teams and country offices abroad. There is a strong focus on program level evaluations (with 112 impact evaluations ongoing and planned) and about an equal number of reviews and other studies are programmed".

SADEV heeft evenmin impact studies uitgevoerd, zij het niet op grond van beleidsoverwegingen maar “basically because they do not have the human resources in-house to undertake such studies”. The M&EU (Sida) heeft voor 2012 beleids ‘reviews’ geprogrammeerd over vormen en methoden van samenwerking op de terreinen van ‘gender’ (dialogues), results management en het meten van successen in conflictsituaties. “It does not carry out impact studies. Instead, Sida has programmed case studies with a long-term perspective on aid cooperation and its effects in three Asian countries”.

Wat betreft het evaluatie-onderdeel van het betreffende Deense ministerie is vastgesteld dat voor 2012 policy reviews van landen-programma’s en landen-gerichte samenwerkings-strategieën geprogrammeerd zijn die als thema’s het trekken van lessen en het ondersteunen van landbouw- en visserij-gerichte activiteiten betreffen. “It plans special studies on particular programs and funds. The department engages in impact studies mainly in the context of joint donor evaluations (budget support to Zambia). Peer reviews of multilateral agencies are also programmed as part of joint donor evaluation efforts”.

Waar het de decentrale evaluatiepraktijk in Zweden en Nederland betreft, blijkt dat die voornamelijk gericht is op project en programma reviews.

Met betrekking tot de vraag of ook *gezamenlijke evaluaties* door de evaluatie-units geprogrammeerd en geëntameerd worden, is het volgende naar voren te brengen. Denemarken en Zweden hebben een beleid en een praktijk waarin grote nadruk wordt gelegd op gezamenlijk evaluaties programmering, maar ook het betrekken van de partnerlanden is een prioriteit. In het mandaat van IOB is het belang van het gezamenlijk uitvoeren van evaluaties opgenomen. Dit is terug te vinden in de programmering. IOB is actief in het gezamenlijk uitvoeren van evaluaties en het betrekken van de partnerlanden of partners die worden geëvalueerd. In feite geldt dit ook voor de centrale evaluatiediensten binnen het ministerie in Denemarken en in Zweden.

Op decentraal niveau bij Buitenlandse Zaken gebeurt gezamenlijk evalueren sporadisch. Voor SADEV is het onderdeel van hun beleid. Het Verenigd Koninkrijk heeft op politiek niveau het belang van ‘joint donor evaluations’ onderkend en benoemd, maar sinds de oprichting van ICAI is dit niet langer meer prioritair.

4.3.4 Enkele vragen voor de toekomst

Onze analyse beoogt niet een ‘endgultige’ studie van evaluatieprogrammering (bij Buitenlandse Zaken/IOB te zijn). Het ging ons er vooral om te kijken hoe in enkele relevante andere landen geprogrammeerd wordt, wat geprogrammeerd wordt en waarom dat gebeurt zoals het gebeurt. Tegen die achtergrond ronden wij onze analyse af met enkele vragen waarvan wij hopen en verwachten dat die in IOB verband en breder op de agenda komen.

✓ *Betrokkenheid van stakeholders bij de programmering*

IOB is in vergelijking met enkele andere evaluatiediensten in de door ons bestudeerde buitenland buitengemeen spaarzaam in het betrekken van maatschappelijke stakeholders (waaronder de politiek) bij de onderzoeksprogrammering. Eerder is wel eens geweest op het ‘Hagiocentrische karakter’ van het onderzoek. Waarom blijven andere ‘stakeholders’, zoals maatschappelijke bewegingen, belangengroepen, politici méer buiten beschouwing dan in enkele van de bezochte andere landen? Aangetekend zij dat wij tegelijkertijd van een Kamerlid vernamen dat betrokkenheid van IOB met de Kamer niet nodig of wenselijk wordt geacht, omdat, als de Kamer contact wenselijk vindt, dat altijd via de Minister mogelijk is.

Naar ons oordeel ligt hier een thema om verder over na te denken, ook nu er mogelijkwijze in verband met de uitbreiding van taken van het beleidsveld met buitenlandse handel, nieuwe stakeholders in beeld komen. IOB heeft gesprekken met diverse beleidsmedewerkers om te horen welke thema’s geprogrammeerd dienen te worden, om de verantwoordelijkheid voor het uitvoeren van evaluaties in opdracht van budgethouders te

benadrukken en het belang van 'quality at entry' te waarborgen. In enkele andere landen worden vooral andere 'stakeholders' betrokken bij het programmeringsproces.

✓ *Selectiviteit in en van het onderzoeksprogramma in plaats van volledige dekkendheid*

In plaats van een dekkende benadering door IOB te wensen, achten we het denkbaar dat meer selectiviteit te overwegen valt. Immers, waarom 'dekkendheid' ten aanzien van bepaalde maatregelen, als uit 'repositories' van *door anderen* verricht onderzoek snel en doeltreffend kennis is af te leiden die in belangrijke, zij het wellicht niet volledige mate, antwoord geeft op de gestelde vragen. Dan wordt 'afdekking' gerealiseerd door uit andere bronnen te putten dan eigen IOB onderzoek. Ook vinden wij het te overwegen om op grond van strategische overwegingen niet steeds voor dekkendheid te opteren, maar voor een zekere selectiviteit (waarbij 'tijdigheid' van onderzoek met het oog op de beleidscyclus een belangrijk onderwerp zou moeten zijn). Ook zou het te overwegen zijn voort te gaan op de weg van decentrale evaluaties waarbij wel een 'toetsmoment' van IOB blijft bestaan ten aanzien van de kwaliteit van de methodologie en data.

✓ *Programmering in relatie tot gebruik en bruikbaarheid van IOB onderzoeken*

Het PvA heeft met enige verbazing moeten vaststellen dat geen informatie beschikbaar is over de mate waarin de huidige wijze van programmering meer of minder bijdraagt aan de beide functies van evalueren: leren en verantwoorden. Gezien de redelijk grote veranderingen in wijze van programmeren in twee van de drie bezochte landen en gezien de constatering dat programmering vooral een proces van goedwillende mensen met al dan niet 'weary eyes' is, bevelen wij aan de uitgangspunten achter de huidige programmering onder de loep te nemen en daarbij met name te letten op de rol die programmering kan spelen bij het vergroten van de bruikbaarheid en het gebruik van IOB-evaluaties. Dit temeer omdat door de departementale herpositionering van het terrein van IOB vermoedelijk ook nieuwe vraagvelden en -stellingen zich aandienen bij een, opnieuw zeer vermoedelijk, niet groter wordende financiële ruimte voor de Inspectie.

✓ *Meer gezamenlijk onderzoek programmeren op top prioriteiten*

Het veld van IOB werd en wordt steeds meer interdepartementaal; de thema's daardoor eerder meer dan minder complex. Daarnaast roept de overlap van vraagstellingen op het gebied van topprioriteiten van het rijk met de prioriteiten van andere donoren de vraag op of meer gezamenlijke evaluaties met andere landen en instituten niet wenselijk zou zijn. Denemarken en Zweden hebben een relatief hoog aantal gezamenlijke evaluaties ten opzichte van het totaal. IOB heeft als trekker en deelnemer deelgenomen aan de meeste belangrijke gezamenlijke evaluaties, zoals de evaluatie van de Paris Declaration, Budget support, vredesopbouw DRC, schuldsanering Nigeria en DRC en de evaluatie van Zuid-Soedan. IOB onderneemt dus al in serieuze mate gezamenlijke evaluaties, maar de vraag kan gesteld of dat niet meer zou moeten gebeuren uit het oogpunt van doelmatigheid. Tevens is de vraag gerechtvaardigd naar de kwalitatieve meerwaarde van gezamenlijke evaluaties, een vraag die ook gesteld kan worden bij 'vertical alignment' respectievelijk de 'dakpansgewijze' programmering. Beantwoording van deze vraag vereist nader onderzoek. Het PvA acht het raadzaam een dergelijk onderzoek te doen uitvoeren.

5. 'Theories of change' en Interventietheorieën in evaluaties

In recente IOB-evaluaties¹⁴ wordt regelmatig gerefereerd aan de aan- of afwezigheid (en de kwaliteit van) *theories of change*, interventielogica's e.d. die worden (of zouden moeten worden) toegepast in de activiteiten of *interventies* (projecten, programma's en *policies*) die in die evaluaties aan bod komen. Het begrip veranderingstheorie komt voort uit de DAC-teksten over evalueren van de relevantie van projecten over conflict, vrede en ontwikkelingssamenwerking. Hiermee verband houdende termen die men in de literatuur kan tegenkomen zijn: programma-, beleids- en interventietheorie (soms ook: - 'logica', 'model', etc.). Op het

14 Zie onder de voor dit tweede advies onderzochte studies: IOB 331, 336, 362 en gezamenlijke evaluaties inzake DR Congo (Channel research, Ohain/BE 2011) en Zuid Sudan (ITAD, UK 2010).

hoogste niveau gaat het dus om beleidsevaluaties,¹⁵ waarin ‘beleidstheorie’ een belangrijke rol speelt: de relevante causale relaties en/of veronderstelde probleem-oorzaak-gevolg relaties en doel-middel relaties.¹⁶ IOB spreekt van een ‘beleidstheorie’ naast een ‘interventiologica’, maar laat ook de eerste de laatste omvatten.¹⁷ Dat draagt niet bij tot helderheid over betekenis en reikwijdte van deze termen. Het is van belang enige ordening en duiding aan te brengen in deze verzameling samenhangende, deels overlappende termen. In sommige van de evaluatiestudies van IOB wordt de afwezigheid van een adequate *theory of change* of beleidstheorie kennelijk negatief beoordeeld, wat de wenselijkheid van meer helderheid onderstreept. Hieronder volgen enkele gedachten van het Panel van Advies over (het gebruik van) deze termen.¹⁸

Een interventie is een samenhangend geheel van activiteiten die wordt ondernomen teneinde een bestaande situatie te veranderen in de richting van een gewenste situatie. Het lijkt helder om een weergave van een interventie in samenhang met zijn effecten in de gehele resultaatketen aan te duiden met de term ‘interventie logica’. Die geeft dan een systematische aanduiding van de samenhang tussen (ontwikkelings-) activiteiten en de verwachte/gerealiseerde effecten. De term ‘Intervention logic’ kan in principe op alle soorten en niveaus van interventies worden toegepast. Termen als *logical frameworks* (of ‘logframes’, ‘logic models’) hebben dan betrekking op methodieken of ‘formats’ voor de weergave van interventielogica’s.

Naast logica of ‘logic’ komen ook termen als ‘policy-’, en ‘program(me) theory’ veelvuldig voor; we kunnen ze generaliseren tot de categorie interventietheorie.¹⁹ Ze bestaan uit “set(s) of assumptions about how and why activities will bring about effects”, waarbij – en daar moeten we expliciet op terugkomen – aandacht is voor daarbij relevante “*behavioural and social mechanisms*” die aan beleid (c.q. interventie) ten grondslag liggen, inclusief de werking daarvan.²⁰ Hanteren van de term ‘theorie’ impliceert het analyseren van *hoe* die mechanismen werken en niet alleen een opsomming daarvan. Het verdient aanbeveling om ‘logics’ van ‘theories’ te onderscheiden en de laatste term te reserveren voor weergaven van waar interventies op gebaseerd zijn met *niet alleen* hun doelen en effecten, maar *ook* met de (veelal contextueel gedifferentieerde) reactiemechanismen en causale samenhangen, die de interventie koppelt aan die effecten (en neveneffecten). Interventietheorieën behandelen dan niet alleen het *wat* van de interventie, maar ook het *hoe* en zelfs (tot op zekere diepte) het *waarom* van de veranderingen, die de al dan niet beoogde resultaten genereren, inclusief de aandacht die wordt geschonken aan relevante contextuele variabelen en mechanismes.²¹

Naast de term ‘interventietheorie’ komt ook het begrip *theory of change* voor. OECD/DAC verstaat daaronder: de veronderstellingen waarop de interventie is gebaseerd inclusief de relaties tussen de interventie en de factoren die het probleem voeden waarop de interventie is gericht in de betreffende context. Dit komt goed overeen met wat elders door OECD een interventietheorie is genoemd en is in lijn met de wetenschappelijke literatuur over dit begrippenkader. De conclusie is dat de begrippen “interventietheorie” en “veranderingstheorie” onderling verwisselbaar zijn. Het gebruik van slechts een van beide in de evaluatiemethodiek zou de conceptuele transparantie bevorderen.

15 Zie RPE 2006. Regeling Prestatiegegevens en Evaluatieonderzoek Rijksoverheid, 2006.

16 Zie RPE 2002. Regeling Prestatiegegevens en Evaluatieonderzoek Rijksoverheid, 2002. Overigens komt de term “beleidstheorie” in RPE 2006 niet langer voor.

17 IOB (2009) *Evaluatiebeleid en richtlijnen voor evaluaties*. pp. 29 en 34.

18 Voor een uitgebreider versie wordt verwezen naar de discussienota J.B. Opschoor (2012): “Theories of Change” en Interventie theorieën in Evaluaties van Ontwikkelingssamenwerking, die beschikbaar is op het Secretariaat van het PvA.

19 Het begrip ‘theorie’ komt in de evaluatieliteratuur nog op een andere wijze naar voren, en wel bij de uitwerking van wat “evaluatieonderzoek” is; zie hiervoor de discussienota waaraan eerder werd gerefereerd.

20 Leeuw, F. and J. Vaessen 2009. *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation*. NONIE, WB.

21 Zie ook Astbury, B. and F. Leeuw 2010. *Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation*. Am. Journal of Evaluation 2010 (Vol 31 no.3: 363-381).

Wat de meer substantiële aspecten betreft, kan men zich afvragen in hoeverre bij interventies een adequate theoretische basis *kan worden verwacht*. In IOB-evaluaties komt het (recentelijk) bovendien herhaaldelijk voor, dat over de in de studie beschouwde interventies wordt opgemerkt dat ze niet of te weinig zouden zijn gerelateerd aan een gearticuleerde adequate beleidstheorie of *theory of change*. De vraag naar een adequate *theory of change* (of interventietheorie) lijkt aldus te evolueren tot een evaluatiecriterium. De vraag is, hoe ver dit moet, of kan, gaan.

Vanzelfsprekend dienen ontwikkelaars zich bij het ontwerpen van hun interventies bewust te zijn van aanwezige kennis en kunde relevant voor veranderingsprocessen van het soort dat zij overwegen, en ook van de context waarin die interventies gaan plaatsvinden. Dat roept de vraag op, hoe diep en breed de scope van de theorie zou moeten zijn. Vaak blijkt echter het aanbod van voor interventies of beleid relevante kennis kwalitatief en kwantitatief onvoldoende te stroken met de behoefte. Ook waar wel sprake is van kennis, kan het zijn dat die onvoldoende eenduidig is (of zelfs contradictoir). Niet zonder grond merkt WRR (2010) op: “Causale schema’s schieten tekort. Veel interventies werken indirect en op termijn, en het is hooguit mogelijk een vermoeden te hebben over hun werking [...] Complexe interventies in maatschappelijke processen goed evalueren is altijd lastig, zo niet onmogelijk. Dat is uiteraard geen argument om het niet zo goed mogelijk te (proberen te) doen. Het is wel een goede reden om niet te hoog van de toren te blazen.” Voor wat betreft de diepgang die haar voor ogen staat bij veranderings- of beleidstheorieën, gaf de eerder genoemde Weiss (1997) ooit een nuttige relativering: ze acht het woord ‘model’ beter dan ‘theory’: “what is asked for is a set of beliefs or assumptions ...hypotheses on what people build their programmes and actions on”. Ook Leeuw en Vaessen (2009) geven als empirische bases voor evaluaties niet alleen *logframes* en literatuur over mechanismen en processen, maar ook: ‘inzichten’ en ‘verwachtingen’ van actoren en stakeholders. In dit verband wijzen zij naast de interventietheorie op de optie van de *causal story* – een ‘verhaal’ over verwachtingen omtrent de doorwerking van de interventie, waarbij vooral de plausibiliteit een toetssteen is (zoals overigens ook in RPE 2006 wordt aanvaard).

Als de evaluatiepraktijk zich steeds meer wenst te baseren op verwachtingen inzake de fundering van interventies in visies op hoe en waarom zij zouden kunnen bijdragen aan de realisatie van gewenste veranderingen, dan kan dat vaak niet gaan om gedetailleerde en harde ‘theorie’, maar om gearticuleerde verwachtingen die plausibel gemaakt wordt in een helder ‘verhaal’. Op een (groeiend) aantal onderdelen zal dat verhaal kunnen worden versterkt met harde hypothesen en feiten, en verdiept door in te gaan op de meest relevante werkzame mechanismes. In elk geval zou wel al bij de toekenning van steun aan interventies moeten zijn gevraagd om te verantwoorden hoe met bestaande kennis en kunde is omgegaan. Maar als van de aanwezigheid van een interventietheorie een criterium wordt gemaakt, dient wel te worden vastgesteld of dat legitiem is in het kader van de *terms of reference* die donor en ontwikkelingsorganisatie zijn overeengekomen, en dienen de evaluatoren ook een gefundeerd beeld te hebben van wat een haalbare en geloofwaardige *theory of change zou kunnen omvatten*.

6. Slotbeschouwing

Tot slot wil het Panel van Advies in concluderende zin reflecteren op een aantal belangrijke punten in dit advies aan de orde zijn gekomen. Tevens wordt in dit slotstuk gekeken naar de mate waarin het panel invulling heeft gegeven aan zijn mandaat en welke onderzoeksvragen in de volgende periode aan de orde zullen zijn.

Hoewel het PvA in het algemeen een positief oordeel heeft over de bruikbaarheid van de bestudeerde rapporten, is dit bij de ene soort evaluatie makkelijker te realiseren dan bij de andere. Bij synthesestudies bijvoorbeeld, kan de bruikbaarheid worden verhoogd door de kwaliteit in deelstudies strenger te bewaken en meer uniformiteit aan te brengen in de gehanteerde begrippenkaders en methoden. Daarmee wordt een betere generalisering mogelijk. Het gebrek aan baseline studies blijft een methodologische handicap voor impact evaluaties. Hierop werd ook al in het eerste advies van het PvA gewezen. Deels kan dit probleem

worden ondervangen door een adequate context analyse vooraf. Een gerelateerd punt van aandacht is dat in de operationalisering soms geen duidelijk verschil wordt gemaakt tussen impact en outcome. In concreto is het van belang dat meer kritische reflectie plaatsvindt op de operationaliseerbaarheid en meetbaarheid van 'outcome' en 'impact' in de verschillende aan de orde zijnde vormen van interventie (c.q. beleid), en op wat dit impliceert voor de opzet en programmering van evaluatieonderzoek. Bij het nastreven van duurzaamheid speelt het probleem dat de beperkte tijdsaders van de interventies de toetsing op duurzaamheid vaak in feite onmogelijk maken. De vraag is waarom IOB in zulke gevallen de evaluaties toch met ambities inzake impact en duurzaamheid optuigt. Dit komt de bruikbaarheid van het advies niet ten goede, omdat verwachtingen worden gewekt die niet waargemaakt kunnen worden.

De ene evaluatie is op grond van de onderzoeksvragen en de ToR van direct belang voor een kleinere groep stakeholders dan de andere. Zo is de bruikbaarheid van de evaluatie Twinning faciliteit Suriname-Nederland in het algemeen beperkt, vanwege het beperkte beleidskader en het specifieke format van twinning. Sommige van de bevindingen kunnen echter wel degelijk een rol spelen in een breder kader. Zo zijn de bevinding van het belang van de tijdsfactor in de effectieve samenwerking tussen maatschappelijke organisaties en de wijze waarop in de evaluatie het begrip maatschappelijk middenveld werd geoperationaliseerd relevant voor het debat over capaciteitsontwikkeling.

Er is in een enkel geval sprake van rapportage die een breder terrein bestrijkt dan in de ToR is neergelegd. Het kan zijn, dat dat vooral daar speelt waar al de ToR tot op zekere hoogte een 'negotiated' karakter draagt en de evaluatoren toch andere (niet in de ToR verankerde) verwachtingen trachten te honoreren, maar geëigend is dit niet. Indien dit punt zich weer dreigt voor te doen, is het van belang dat de onafhankelijke positie van IOB inzake ToRs prevaleert. Een uitbreiding van de rapportage naar onderwerpen die strikt genomen buiten de ToR vallen kan belangrijk zijn (bijvoorbeeld de kwestie van armoedebestrijding in de evaluatie Twinning faciliteit Suriname-Nederland), maar is alleen acceptabel als de kwaliteit van de verzamelde data dat toelaat. Als de empirische basis voor verderstreckende antwoorden niet aanwezig is, krijgen dergelijke uitweidingen een ongefundeerd en willekeurig karakter en kan een en ander ook leiden tot additionele vragen die de evaluatie heeft opgeroepen.

Ten aanzien van de evaluatieprogrammering kan geconcludeerd worden dat toepassing van een vergelijkend perspectief op de wijze waarop de programmering in verschillende landen tot stand komt, laat zien hoe deze samenhangt met de localisatie van de evaluatiefunctie en de daaraan gerelateerde verantwoordelijkheid voor de besluitvorming over de programmering. Het punt van de optimale balans tussen de accountability- en de leerfunctie van evaluaties wordt in het IOB programmeringsproces niet geproblematiseerd. Aan de eerste functie lijkt een groter gewicht te worden toegekend dan de tweede, zoals tot uiting komt in het streven de programmering 'dekkend' te laten zijn in plaats van meer selectief te werk te gaan. Externe stakeholders en het maatschappelijk debat spelen niet of nauwelijks een rol in de huidige praktijk van de evaluatieprogrammering. Met de veranderde politieke positionering van het beleidsterrein van Ontwikkelingssamenwerking (koppeling met buitenlandse handel) komen nieuwe stakeholders in beeld en verdient dit punt aandacht. Ook is er uit het oogpunt van doelmatigheid wat te zeggen voor het meer dan tot dusverre programmeren van gezamenlijke en samengestelde (dakpansgewijs) evaluaties. Er is het panel echter geen onderzoek bekend naar de meerwaarde in termen van kwaliteit van dit soort evaluaties en de mogelijke 'trade-off' tussen doelmatigheid en kwaliteit. Het meer programmeren van gezamenlijke en samengestelde evaluaties zou van invloed kunnen zijn op de balans tussen leren en verantwoorden in de evaluatieprogrammering, maar zolang die balans geen punt van expliciete overweging is in de totstandkoming van de programmering valt ook hiervan weinig te zeggen.

Voor wat betreft de methodologieontwikkeling in evaluaties zelf is een nadere reflectie op verwachtingen inzake gehanteerde 'theories of change' en hun bruikbaarheid bij het beoordelen van interventies van belang.

Literatuurlijst

- Andersen O.W. and Broegaard, E. (2012) *The Political Economy of joint Evaluation*. Evaluations, 18 (1), 47-59
- Astbury, B. and F. Leeuw (2010) *Unpacking Black Boxes: Mechanisms and Theory Building in Evaluation*. Am. Journal of Evaluation, Vol. 31 no.3:363-381
- Bennett, J., S. Pantuliano, W. Fenton, A. Vaux, C. Barnett, and Brusset, E. (2010) *Aiding the Peace. A Multi-Donor Evaluation of Support to Conflict Prevention and Peacebuilding Activities in Southern Sudan 2005 – 2010*. ITAD Ltd., United Kingdom.
- Channel Research (2011) *Misschien Vrede – gezamenlijke evaluatie van conflictpreventie en vredesopbouw in de Democratische Republiek Congo, Syntheseverslag*. Brussel: FOD Buitenlandse Zaken, Buitenlandse Handel en Ontwikkelingssamenwerking.
- De Ruijter, A. en De Gaay Forman, B. (2008) *Van overkant tot overkant*
- Department for International Development (2009) *Building the evidence to reduce poverty. The UK's policy on evaluation for international development*. Londen/Glasgow: DFID
- Department for International Development (2009) *Independent evaluation in DFID*. Annual report 2009. Londen/Glasgow: DFID
- Development Assistance Committee (2011) *Peer Review of Denmark 2011*. Paris: OECD
- Development Assistance Committee (2011) *Peer Review of Sweden 2009*. Paris: OECD
- Development Assistance Committee (2011) *Peer Review of The Netherlands 2011*. Paris: OECD
- Development Assistance Committee (2011) *Peer Review of the United Kingdom 2010*. Paris: OECD
- Gils, van, G.H.C., Leeuw, F.L., Gildemyn, M, Inberg, L. (2010) *Leren van evalueren. Onderzoek naar het gebruik van evaluatie onderzoek bij het ministerie van Buitenlandse Zaken*. Den Haag
- Helmsing, A.H.J. (2012) *Evaluation and its programming: a focus on the rear view mirror? Rapport gemaakt in opdracht van het Panel van Advies IOB, Den Haag, 2012*
- Hoesel, P.H.M van, *Programmering van beleidsonderzoek*. Dissertatie RU Leiden, 1985
- Independent Commission for Aid Impact (2011) *Corporate Plan*. London: ICAI
- Independent Commission for Aid Impact (2011) *ICAI's Approach to Effectiveness and Value for Money*. Londen: ICAI Report Nr 1. London: ICAI
- Independent Commission for Aid Impact (2011) *Work Plan*. London: ICAI
- Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (2011) *Evaluation of Dutch support to Capacity Development. The case of the Netherlands Institute for Multiparty Democracy (NIMD)*, no. 331
- Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (2011) *Facilitating Resourcefulness. Synthesis report of the Evaluation of Dutch support to Capacity Development*, no. 336
- Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (2011) *Table rase – et après? Evaluation de l'Allègement de la Dette en République Démocratique du Congo 2003-2010*, no. 347
- Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (2011) *Education matters: Policy review of the Dutch contribution to basic education 1999–2009*, no. 353
- Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (2011) *Assisting Earthquake victims: Evaluation of Dutch Cooperating aid agencies (SHO) Support to Haiti in 2010*, no. 358
- Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (2011) *Evaluatie van de Twinningfaciliteit Suriname-Nederland*, no 361

Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (2011) *De Methodische kwaliteit van Programma-evaluaties in het Medefinancieringsstelsel-I 2007-2010*, no. 362

Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (IOB) (2012) *Jaarbericht IOB 2011*. Powerpoint presentatie

Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (IOB) (2011) *Evaluatieprogrammering Begroting 2012*. Ministerie van Buitenlandse Zaken: Den Haag

Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (IOB) (2009) *Evaluatiebeleid en richtlijnen voor evaluaties*. Ministerie van Buitenlandse Zaken: Den Haag

Lakatos (1978). *The Methodology of Scientific Research Programmes*: Philosophical Papers Volume 1. Cambridge: Cambridge University Press

Leeuw, F. and J. Vaessen (2009) *Impact Evaluations and Development: NONIE Guidance on Impact Evaluation*. NONIE, WB.

Lindblom, Ch. (1959) *The science of muddling through*. Originally published in the Public Administration review, 19, 79-88. Reprinted in Gawthrop, L.C. 1970. The administrative process and democratic theory. Boston: Houghton Mifflin Company.

Manning, R. (2009) *The Quality of DFID's Evaluation and Evaluation Management Systems: How do they compare with other agencies?* Consultant report to the Independent Advisory Committee on development Impact. Londen: DFID

Ministerie van Financiën (2002) *Regeling Prestatiegegevens en Evaluatieonderzoek Rijksoverheid, 2002*.

Ministerie van Financiën (2006) *Regeling Periodiek Evaluatieonderzoek en Beleidsinformatie*. Staatscourant 26 april 2006, nr. 83

Ministry of Foreign Affairs of Denmark/ DANIDA (2012) *DANIDA Evaluation Guidelines*. Copenhagen

OECD/DAC (2010) *Development evaluation resources and systems – A study of network members*. The DAC Network on Development Evaluation. Paris: OECD.

Panel van Advies (2011) *Eerste Advies mei 2011 naar het gebruik van IOB evaluaties*

Patton, M.Q. (2008) *Utilization-focused evaluation*. Editie 4. Londen: Sage

Perrin, B. (2009) *Review of the Quality of DFID Evaluations. A Delicate Balancing Act*. Consultant report to the Independent Advisory Committee on Development Impact. London: DFID

Riddell, R.C. (2009) *The quality of DFID's evaluation reports and assurance systems*.

Riddell, R.C. (2009) *The quality of DFID's evaluation reports and assurance systems*.

SADEV (2011) *The theme for SADEV's evaluation programme 2012 is adaption*. Internal document SADEV. Karlstad: SADEV

Sida (2012) *Sida. Strategic Evaluation Plan 2012*. The Unit for Monitoring & Evaluation (UTV), department for Organisational Development (VU). Stockholm: Sida

Sida (2012) *Sida's Evaluation Guidelines 2012*. The Unit for Monitoring & Evaluation (UTV), department for Organisational Development (VU). Stockholm: Sida

Tweede Kamer, vergaderjaar 2007-2008, 31 200 V, nr. 136.

Tweede Kamer, vergaderjaar 2008-2009, 29 237, nr. 83.

Tweede Kamer, vergaderjaar 2008-2009, 29 237, nr. 85.

Tweede Kamer, vergaderjaar 2008-2009, 29 237, nr. 99.

Weiss, C. (1997) *Bringing theory-based evaluation within our means. Harvard Project on Schooling and Children.* Harvard Graduate School of Education.