IOB Evaluatie

Methodische kwaliteit van Programma-evaluaties in het Medefinancieringsstelsel-I 2007 – 2010

Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie

Meta-evaluatie

De term meta-evaluatie werd in 1969 geïntroduceerd door Michael Scriven, een gezaghebbende auteur op het gebied van evaluatie, als aanduiding voor de evaluatie van een evaluatieonderzoek, een evaluatiesysteem of een evaluatie-instrument.

Over de inhoud van het begrip kan verwarring ontstaan doordat het nogal eens – ten onrechte – wordt gehanteerd voor de compilatie van een aantal (deel)evaluaties tot een samenvattende synthesestudie.

In dit rapport wordt meta-evaluatie uitsluitend gebruikt in de oorspronkelijke betekenis van evaluatie van evaluatieonderzoek.

Voorwoord

De Nederlandse overheid besteedt een deel van haar internationale ontwikkelingssamenwerking via Nederlandse maatschappelijke organisaties. Deze zogenaamde medefinancieringsorganisaties onderhouden een breed netwerk van samenwerkingsrelaties met (inter)nationale en Zuidelijke partners en ontplooien een scala aan activiteiten op de terreinen van directe armoedebestrijding, maatschappijopbouw en capaciteitsontwikkeling, en beleidsbeïnvloeding.

Binnen het kader van het eerste Medefinancieringsstelsel (MFS-I) voerden de organisaties onder eigen verantwoordelijkheid evaluatieonderzoek uit. Deze evaluaties waren bedoeld om de eigen beleidsvorming te oriënteren, en dienden daarnaast ter verantwoording van de bestedingen. De Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (IOB) van het Ministerie van Buitenlandse Zaken heeft vervolgens de uitgevoerde evaluaties getoetst aan de methodische kwaliteitscriteria validiteit, betrouwbaarheid en bruikbaarheid. Deze criteria zijn verder geoperationaliseerd en toegepast op een totaal van 72 programma-evaluaties van elf medefinancieringsorganisaties. Hiermee is ruim € 780 miljoen aan subsidie-uitgaven geëvalueerd, hetgeen overeenkomt met 47% van de MFS-subsidietoekenning.

De bevindingen van deze meta-evaluatie laten zien dat er weliswaar verbeteringen zijn opgetreden in de evaluatiekwaliteit van de MFS-organisaties, maar dat er op kritische onderdelen ook nog steeds belangrijke tekortkomingen worden gesignaleerd. Vooral op het gebied van de interventielogica, de definiëring van resultaatniveaus en het hanteren van specifieke, meetbare indicatoren blijft er nog veel te wensen over. Speciale aandacht wordt ook gevraagd voor de methoden van resultaatmeting, waarbij de representativiteit van de onderzochte activiteiten en de mogelijkheden voor de generaliseerbaarheid van de bevindingen nog ernstig tekortschieten.

Dit onderzoek is uitgevoerd door IOB inspecteur Dick van der Hoek en maakt gebruik van de door IOB ontwikkelde beoordelingslijst om de opzet en uitvoering van programma-evaluaties binnen het Medefinancieringsstelsel systematisch te toetsen. Het rapport is van deskundig commentaar voorzien door Dr. Jos Vaessen, evaluatie-expert bij UNICEF.

IOB is de betrokken MFS-organisaties erkentelijk voor hun constructieve medewerking aan dit onderzoek. Het frequente overleg dat plaatsvond voorafgaand aan de vaststelling van de beoordelingen heeft ongetwijfeld bijgedragen aan de versterking van het leerproces binnen de betrokken organisaties.
IOB is eindverantwoordelijk voor de inhoud van deze evaluatie.

Prof. dr. Ruerd Ruben
Directeur
Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie
Ministerie van Buitenlandse Zaken

Inhoudsopgave

Voorwoord 3
Tabellen en figuren 7
Lijst van afkortingen 8

1 Hoofdbevindingen 10

2 Inleiding 14
2.1 Achtergrond 15
2.1.1 Medefinancieringsprogramma-breed 16
2.1.2 Medefinancieringsstelsel 17
2.2 Doelstelling 18
2.3 Probleemstelling 19
2.4 Rapportopzet 19

3 Onderzoekobject 22
3.1 MFS-I organisaties 23
3.2 Soorten evaluaties 24
3.2.1 Projectevaluaties 24
3.2.2 Programma-evaluaties 26
3.3 'Evaluatiedekking' 29
3.4 Verantwoordingsvormen 31

4 Onderzoekaanpak 32
4.1 Regeling Prestatiegegevens en Evaluatieonderzoek (RPE) 33
4.2 Beoordelingscriteria 34
4.2.1 Validiteit 36
4.2.2 Betrouwbaarheid 38
4.2.3 Bruikbaarheid 40
4.3 Waarderingsmethode 41
4.4 Beoordelingsprocedure 44
4.4.1 Aanlevering van studies 44
4.4.2 Concept-beoordeling 46
4.4.3 Bespreking en vaststelling 47

5 Onderzoekresultaten 48
5.1 Evaluatiedekking 49
5.2 Algemene PE-kenmerken 54
5.3 Evaluatiekwaliteit 56
5.3.1 Validiteit 56
5.3.2 Betrouwbaarheid 59
5.3.3 Bruikbaarheid 62
5.4 Lerend vermogen 64
5.4.1 Evaluatiefunctiebeoordeling – ex ante vs. ex post 64
5.4.2 Leereffecten 66

Bijlagen **68**
Bijlage 1 Over IOB 69
Bijlage 2 Algemene kenmerken Programma-evaluaties 71
Bijlage 3 Beoordelingslijst Programma-evaluaties 81
Bijlage 4 Conversie van MFP-b naar MFS-I beoordelingsscores 87
Bijlage 4a Beoordelingslijst Programma-evaluaties 90
Bijlage 4b Beoordelingslijst Programma-evaluaties 93
Bijlage 5 Scorekaarten 96
Bijlage 6 Bronnen 114

Tabellen en figuren

Tabellen

3–1	Subsidietoekenning grote MFS-I organisaties	23
3–2	Aantallen Programma-evaluaties	27
5–1	Evaluatiedekking Programma-evaluaties	50
5–2	Scores Organisatietoets: Lerend vermogen	65

Figuren

1–1	Verdeling Beoordelingsscores Programma-evaluaties	11
4–1	Beoordelingspiramide	35
4–2	Vuistregels voor scoretoekenning	42
4–3	Aanlevering Programma-evaluatierapporten per kwartaal 2007–2011	45
5–1	Scores Onderzoekvragen	57
5–2	Scores Evaluatiecriteria	57
5–3	Scores Definitie Evaluatieobject	57
5–4	Scores Context Evaluatieobject	57
5–5	Scores Interventielogica	58
5–6	Scores Indicatoren	58
5–7	Scores Gegevensanalyse	59
5–8	Scores Onderbouwing Conclusies	59
5–9	Scores Onderzoekmethoden	60
5–10	Scores Triangulatie	60
5–11	Scores Representativiteit	60
5–12	Scores Beperkingen	60
5–13	Scores Bronmateriaal	61
5–14	Scores Onderzoekers	61
5–15	Scores Verantwoording	62
5–16	Scores Kwaliteitsbewaking	62
5–17	Scores Onderzoekdoelstelling	63
5–18	Scores Toegankelijkheid	63
5–19	Scores Beantwoording Onderzoekvragen	63
5–20	Scores Aanbevelingen	63
5–21	Scorevergelijking MFP-breed – MFS-I Programma-evaluaties van Cordaid, Hivos, Icco en Oxfam Novib	66

Lijst van afkortingen

DSI	Directie Sociale en Institutionele Ontwikkeling
DSO	Directie Sociale Ontwikkeling
DGIS	Directoraat-Generaal Internationale Samenwerking
E&B	Evaluatie & Beleid (MBN-werkgroep)
ERG	Evaluatie Referentie Groep
GOM	Gemeenschappelijk Overleg Medefinanciering
IBO	Interdepartementaal Beleidsonderzoek
IOB	Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie
MBN	Medefinancieringsprogramma Breed Netwerk
M&E	Monitoring en Evaluatie
MF	Medefinanciering
MFO	Medefinancieringsorganisatie
MFP	Medefinancieringsprogramma
MFP-b	Medefinancieringsprogramma-breed
MFS	Medefinancieringsstelsel
MM	Maatgesneden Monitoring
NGO	Niet-Gouvernementele Organisatie
OS	Ontwikkelingssamenwerking
PE	Programma-evaluatie
PO	Partnerorganisatie
RPE	Regeling Prestatiegegevens en Evaluatieonderzoek (Rijksoverheid)
SMART	*Specific – Measurable – Attainable –Relevant –Time-bound*
TMF	Thematische medefinanciering
ToR	Terms of Reference
VBTB	Van Beleidsbegroting Tot Beleidsverantwoording

1

Hoofdbevindingen

Inleiding

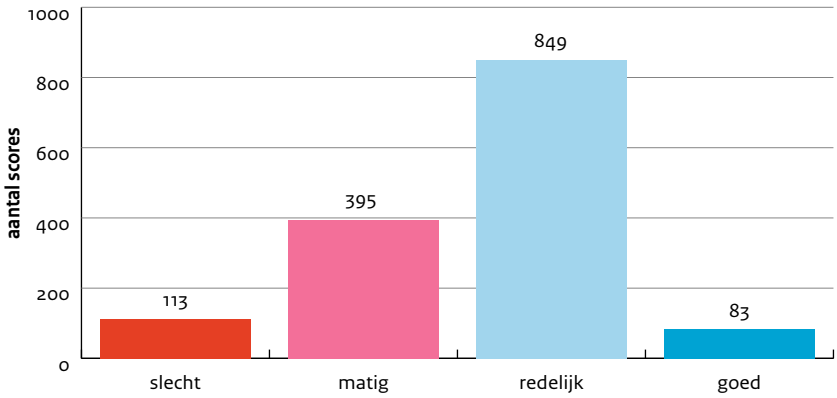
Gedurende de looptijd van het eerste Medefinancieringsstelsel (MFS-I) – 2007 t/m 2010 – heeft de Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (IOB) de methodische kwaliteit van door MFS-organisaties zelf ondernomen evaluatieonderzoek getoetst. Daartoe analyseerde en beoordeelde IOB 72 programma-evaluaties (PE's) van de elf grootste subsidieontvangers. Dergelijke onderzoeken zijn over het algemeen breder van opzet dan evaluaties van individuele projecten en beogen conclusies te trekken op een beleidsniveau dat de individuele samenwerkingsactiviteit overstijgt. De 72 studies bestreken € 756 miljoen aan bestedingen ten laste van medefinancieringssubsidies en vergden zelf in totaal € 6,1 miljoen aan onderzoekuitgaven. De gegevensverzameling voor deze meta-evaluatie werd afgesloten in mei 2011.

De beoordeling is uitgevoerd aan de hand van de drie kwaliteitscriteria ontleend aan de Regeling Prestatiegegevens en Evaluatieonderzoek (RPE) voor de Rijksoverheid: Validiteit, Betrouwbaarheid en Bruikbaarheid. Deze criteria zijn geoperationaliseerd door middel van tien indicatoren, die vervolgens werden uitgesplitst in twintig beoordelingscomponenten. De 72 PE's zijn op elk van deze componenten beoordeeld, waarbij de scores werden vastgelegd op een vierpuntsschaal met waarderingen variërend van slecht via matig en redelijk tot goed. De beoordelingen zijn tot stand gekomen na overleg in twee ronden – een schriftelijke en een mondelinge – tussen enerzijds IOB en anderzijds evaluatievertegenwoordigers en –programmabeheerders van de betrokken MFS-organisaties. Om misverstanden te voorkomen moet de kwaliteit van evaluatieonderzoek wel scherp worden onderscheiden van de kwaliteit van de geëvalueerde activiteiten. Uit de beoordelingen van programma-evaluaties in dit rapport vallen dan ook geen conclusies te trekken over de doelmatigheid en doeltreffendheid van de programma's zelf.

Algemeen

De beoordelingen van de 72 programma-evaluaties op de twintig onderdelen leverde in totaal 1.440 waarderingen op die als volgt over de score-schaal waren verdeeld:

Figuur 1–1 *Verdeling Beoordelingsscores Programma-evaluaties*



Zoals deze verdeling laat zien, is in twee derde van de gevallen (65 %) de waardering positief, dat wil zeggen dat de score *goed* of *redelijk* is, terwijl 35 % (508) van de scores in meer of mindere mate te wensen overlaat. Dit zijn er, *over het geheel genomen*, te veel om de bestudeerde programma-evaluaties te kunnen kwalificeren als een afdoende manier om de betreffende programma-uitgaven te verantwoorden. Daarbij moet wel worden aangetekend dat deze totalen aanzienlijke verschillen tussen de elf organisaties verbergen. Zo zijn er significante verbeteringen geconstateerd bij de grootste subsidieontvangers wier programma-evaluaties al sinds de voorafgaande subsidieperiode (2003 $^{t}/_{m}$ 2006) op hun methodologische kwaliteiten worden beoordeeld. Aan de andere kant kwamen bijna de helft van de 'slechte' en een derde van de 'matige' scores voor bij twee kleinere organisaties die tijdens de afgelopen subsidieperiode voor het eerst met extern toezicht op hun evaluatiekwaliteit in aanraking kwamen.

Validiteit

Uit de waarderingen voor de operationalisering van probleemstellingen in onderzoeksvragen blijkt dat opdrachtgevers van PE's meestal een behoorlijke voorstelling hebben van wat ze uitgezocht willen zien. Daarbij onderschatten ze wel vaak het belang van een expliciete definiëring van de evaluatiecriteria waaraan de interventieresultaten getoetst dienen te worden. Dit laatste geldt ook, zij het in iets mindere mate, voor de heldere afbakening van onderzoeksobjecten en voor duidelijke contextbeschrijvingen.

Echt mis gaat het echter vaak bij de weergave of reconstructie van de interventielogica met de bijbehorende definiëring van resultaatniveaus en specifieke, meetbare indicatoren. Bij deze onderdelen blijft er nog veel te wensen over; vooral omdat de op deze punten geconstateerde tekortkomingen een ondubbelzinnige vaststelling van resultaten ondermijnen.

Betrouwbaarheid

De meest positieve waarderingen onder dit criterium gelden de onafhankelijkheid van de ingeschakelde onderzoekers en de identificatie en verantwoording van de gebruikte menselijke en documentaire gegevensbronnen.

Daarentegen wordt de betrouwbaarheid van PE's aangetast door de nogal eens tekortschietende aandacht voor de representativiteit van de daadwerkelijk onderzochte activiteiten en bijgevolg voor de generaliseerbaarheid van de evaluatiebevindingen. Gezien het dominante gebruik van kwalitatieve evaluatiemethoden zou verder een ruimere en meer expliciete combinatie van verschillende onderzoektechnieken en gegevensbronnen met het oog op de noodzakelijke verificatie van onderzoekbevindingen (triangulatie) voor de hand hebben gelegen.

Bruikbaarheid

Afgezien van het voorgaande zijn er weinig factoren die bruikbaarheid van de beoordeelde PE's beperken. De opdrachtgevers bij de MFS-organisaties blijken over het algemeen goed te spreken over de bruikbaarheid van de opgeleverde evaluatie-uitkomsten.

Samenvattend moet het eindoordeel luiden dat, hoewel er verbeteringen zijn opgetreden in de evaluatiekwaliteit, deze op kritische onderdelen nog teveel tekortschiet om er een

gefundeerd oordeel op te kunnen baseren over de effectiviteit van de geëvalueerde programma's in termen van bereikte resultaten. Om een dergelijk oordeel wel mogelijk te maken zijn in ieder geval (verdere) verbeteringen noodzakelijk in de reconstructie van interventielogica's, de operationalisering daarvan door middel van indicatoren en de representativiteit van steekproeven en casestudies.

2

Inleiding

Dit hoofdstuk schetst in kort bestek de achtergrond en opzet van het IOB-onderzoek naar de kwaliteit van programma-evaluaties[1] uitgevoerd onder het eerste Medefinancieringsstelsel (MFS-I 2007 $^{t}/_{m}$ 2010).

2.1 Achtergrond

Met de invoering van het programmafinancieringsmodel in 1980 deed de zogenoemde programma-evaluatie (PE) haar intrede in het toenmalige Medefinancieringsprogramma (MFP). Deze PE's betroffen doorgaans onderzoeken rond een bepaald thema of in een specifieke regio, waarbij de activiteiten van één van de destijds vier[2] Medefinancieringsorganisaties (MFO's) centraal stonden. Tussen 1980 en 1999 werden er in opdracht van het ministerie van Buitenlandse Zaken en het Gemeenschappelijk Overleg Medefinanciering (GOM) 80 van deze programma-evaluaties ondernomen. Aan de uitvoering ervan namen de MFO's – via het GOM – zelf actief deel, hetgeen tegen het einde van de jaren negentig leidde tot toenemende kritiek op het gebrek aan onafhankelijkheid van de onderzoeken. Bovendien zouden de PE-rapportages in het algemeen onvoldoende resultaatgericht zijn, althans te weinig concrete informatie over daadwerkelijk behaalde resultaten opleveren.

Om aan deze bezwaren tegemoet te komen gingen het GOM en het Directoraat-Generaal Internationale Samenwerking (DGIS) in 1998 over tot de instelling van de onafhankelijke Stuurgroep Evaluatie Medefinancieringsprogramma. Deze stuurgroep, onder voorzitterschap van prof. dr. A. de Ruijter, werd belast met de opzet en uitvoering van de programma-evaluaties, waarvan zij er in haar vierjarige bestaan (1999 – 2002) veertien voltooide (naast een groot aantal kleinere bureaustudies en werkdocumenten). Indachtig de aanleiding tot haar oprichting begon de Stuurgroep de invulling van haar mandaat met een gedetailleerde balans van de stand van zaken rond de PE's. Het resultaat was de *Balansstudie* die onder meer een 'systematisch falen van de effectmetingsfunctie vaststelde[3], waardoor veel van de in de periode 1980 – 1999 uitgevoerde PE's onvoldoende inzicht boden om de doeltreffendheid en doelmatigheid van de onderzochte programma's te beoordelen.[4]

Aan deze laatste bevinding werd veel gewicht toegekend in het kader van het Interdepartementaal Beleidsonderzoek (IBO) van het MFP, dat in 1999 en 2000 werd uitgevoerd door de Werkgroep IBO-MFP onder voorzitterschap van prof. dr. L.B.M. Mennes. Dit onderzoek concludeerde onder andere dat er onvoldoende inzicht te krijgen was in de doeltreffendheid en doelmatigheid van het MFP en dat de minister voor Ontwikkelingssamenwerking (OS) te weinig mogelijkheden had om het programma te sturen.[5] Behalve een duidelijker beleidskader bepleitte het IBO ook een scherper toezicht op de naleving daarvan.

1 Hoewel programmaevaluatie eigenlijk één woord is, wordt het ter wille van de leesbaarheid doorgaans met een koppelteken gespeld: in overeenstemming hiermee is de in het MFS gangbare afkorting: PE.
2 Cebemo (thans: Cordaid), Hivos, ICCO en Novib (thans: Oxfam Novib).
3 Balansstudie (1999): p. 45.
4 Id, p. 6.
5 Brief van de minister voor Ontwikkelingssamenwerking aan de Tweede Kamer d.d. 2 oktober 2000, TK: 2000–2001, 27 433, nr. 1, p. 1.

2.1.1 Medefinancieringsprogramma-breed

In overeenstemming met deze aanbeveling formuleerde de minister voor OS een geheel nieuw raamwerk voor het medefinancieringsprogramma: het Beleidskader MFP-breed.[6] Hoewel de toevoeging van de term 'breed' aan de programmanaam geen betrekking had op de omvang van het programma, breidde het aantal Medefinancieringsorganisaties (MFO's) dat ermee werd gesubsidieerd, zich wel geleidelijk uit. Nadat met ingang van 2000 Foster Parents Plan Nederland (tegenwoordig Plan Nederland) al tot het MFP was toegelaten, werd bij de start van het MFP-breed ook aan de kinderhulporganisatie Terre des Hommes Nederland een medefinancieringssubsidie toegekend. Hiermee steeg het totale aantal MFO's tot zes.

Tijdens de parlementaire behandeling van de nieuwe opzet van het MFP drong de Tweede Kamer bij de minister krachtig aan op maatregelen om de in het IBO-rapport gesignaleerde gebreken te corrigeren. Mede als gevolg hiervan is de rol van IOB in de definitieve versie van het MFP-b beleidskader dat per 1 januari 2003 voor een periode van vier jaar van kracht werd, als volgt verwoord:

IOB inspecteert en toetst de kwaliteit van het geleverde werk, op basis van eigen evaluaties en met gebruikmaking van de uitkomsten van het samenhangende evaluatiesysteem dat de MFO's zelf hanteren.[7]

Aan de invulling van dit mandaat werd door de Inspectie op twee manieren vorm gegeven:

- Door middel van een beperkt aantal 'eigen' IOB-evaluaties van werkwijzen en resultaten van MFO-activiteiten, uitmondend in afzonderlijke evaluatierapporten.
- Door toezicht op de kwaliteit van de evaluaties die de MFO's zelf uitvoeren.[8] Hiertoe werd elke door MFO's individueel of gezamenlijk ondernomen PE achteraf beoordeeld op een aantal methodologische criteria.

De resultaten van dit toezicht op de methodologische kwaliteit van de PE's binnen het MFP-breed werden vastgelegd in het IOB Werkdocument *de Kwaliteit van Programma-evaluaties in het Medefinancieringsprogramma 2003 – 2006* (juni 2006). In deze meta-evaluatie beoordeelde IOB alle 26 PE's die tijdens de subsidieperiode van het MFP-breed waren uitgevoerd. Vier daarvan maakten deel uit van de gezamenlijke meerjarige onderzoekagenda van het MFP-breed Netwerk (MBN).[9] De overige 22 hadden betrekking op evaluatiestudies onderno-

6 De toevoeging *breed* reflecteerde de beperking van het nieuwe programma tot NGO's met een brede, alomvattende benadering van structurele armoedebestrijding met gebruikmaking van de drie samenhangende interventiestrategieën: directe armoedebestrijding, maatschappijopbouw en beleidsbeïnvloeding (TK: 2000-2001, 27433, nr. 2, pp. 5, 11).

7 Herziene versie van het Beleidskader MFP-breed, door de minister voor OS aan de Tweede Kamer aangeboden met haar brief van 22 november 2001, TK: 2001–2002, 27 433, nr. 6, p. 11.

8 Brief van de Staatssecretaris voor Buitenlandse Zaken d.d. 18 november 2002, no. buza 020459, p. 4.

9 In 2004 was het GOM omgedoopt tot het Medefinancieringsprogramma-breed Netwerk, afgekort MBN, waartoe behalve de vier 'oude' MFO's ook Plan Nederland toetrad. Terre des Hommes bleef erbuiten en nam dus ook geen deel aan de gezamenlijke PE's.

men door de vijf individuele MBN-leden.[10] Met de geëvalueerde activiteiten was een totaal bedrag aan programma-uitgaven van € 364 miljoen gemoeid.
De belangrijkste conclusie uit het IOB-onderzoek was dat, ondanks een aantal positieve elementen, de bestudeerde PE's op essentiële onderdelen van zowel validiteit als betrouwbaarheid teveel tekortkomingen vertoonden om te kunnen kwalificeren als adequate verantwoording van de bestedingen.

Tegelijk met het nieuwe MFP-breed voor brede algemene ontwikkelingsorganisaties werd in 2003 een beleidskader geïntroduceerd voor thematisch en/of sectoraal gespecialiseerde organisaties die niet alle drie de strategieën (armoedebestrijding, maatschappijopbouw en beleidsbeïnvloeding) hanteerden, het zogenoemde programma voor thematische medefinanciering (TMF).[11] Vergeleken met het MFP-breed ging het bij het TMF-programma om veel kleinere subsidiebedragen maar om veel grotere aantallen subsidieontvangers: zo werden in 2003 en 2004 aanvragen van 132 organisaties goedgekeurd, terwijl een nog groter aantal werd afgewezen.[12] Deze nieuwe subsidiefaciliteit had dus niet te klagen over belangstelling uit het maatschappelijk middenveld. Desondanks was het programma geen lang leven beschoren.

2.1.2 Medefinancieringsstelsel

Al in het jaar dat het TMF-programma van start ging (2003) besloot de (inmiddels nieuwe) minister voor Ontwikkelingssamenwerking dat er geen inhoudelijke reden bestond om MFP en TMF strikt gescheiden te houden. De samenvoeging zou daarom met ingang van 2007, bij de afloop van de MFP-breed financiering haar beslag kunnen krijgen.[13] Aldus geschiedde, zodat eind 2004 een ontschot MFP-breed/TMF beleidskader kon worden gepresenteerd onder de benaming *Medefinancieringsstelsel* (MFS)[14] 2007–2010.[15] Dit nieuwe stelsel zou plaats bieden aan zowel brede organisaties die verscheidene thema's bestrijken op basis van met elkaar samenhangende interventiestrategieën, als aan kleinere gespecialiseerde organisaties die zich concentreren op een prioritair thema en daarbij minder dan de drie genoemde strategieën hanteren.

Omdat de resultaten van het IOB-toezicht op de methodologische kwaliteit van PE's binnen het MFP-breed hadden laten zien dat er nog veel aan deze evaluaties te verbeteren viel, lag het voor de hand dat het kwaliteitstoezicht van de Inspectie onder het nieuwe stelsel zou

[10] Ten tijde van haar toetreding tot het MFP-b voerde Terre des Hommes nog geen programma-evaluaties uit. Omdat de organisatie de capaciteit daarvoor nog moest ontwikkelen werden gedurende de subsidieperiode 2003–3006 door IOB nog geen formele beoordelingen van evaluatiestudies uitgevoerd.

[11] Brief van de minister voor OS d.d. 6 maart 2002, TK: 2001–2002, 27 433, nr. 7.

[12] Syntheserapport TMF Evaluatie (2006), p. 20.

[13] Beleidsnotitie *Aan elkaar verplicht, Ontwikkelingssamenwerking op weg naar 2015* (oktober 2003), TK: 2003–2004. 29 234, nr. 2, p. 23.

[14] Nadat in 2008 het principebesluit was gevallen het stelsel, zij het in gewijzigde opzet, na 2010 onder dezelfde naam te continueren, ging het subsidiestelsel voor de periode 2003–2006 MFS-I heten. Ter vermijding van verwarring met het dit jaar van start gegane Medefinancieringsstelsel-II (MFS-II) wordt in dit rapport voor het oorspronkelijke stelsel de aanduiding MFS-I gehanteerd.

[15] Bijlage bij brief van de minister voor Ontwikkelingssamenwerking aan de Tweede Kamer d.d. 22 december 2004, TK: 2004–2005, 27 433, nr. 29.

worden voortgezet. Daar het aantal subsidieontvangers onder het MFS-I naar verwachting een veelvoud zou bedragen van de vijf die onder het MFP-breed op hun PE-kwaliteit waren beoordeeld. stond bij voorbaat vast dat het IOB-toezicht niet het hele MFS-I zou kunnen omvatten. Er zou dus een selectie moeten worden gemaakt om met de beschikbare inspectiecapaciteit een zo groot mogelijk deel van het subsidiestelsel te bestrijken. Hiervoor moesten eerst de subsidietoekenningsbeslissingen bekend worden, zodat duidelijk zou zijn welke organisaties hoeveel MFS-I subsidie zouden ontvangen. Dat was pas in september 2006 het geval toen de succesvolle aanvragers te horen kregen welk bedrag in het kader van het MFS 2007–2010 aan hen was toegekend.

Vervolgens werd op voorstel van de Directie Sociale en Institutionele Ontwikkeling (DSI), belast met de coördinatie van het MFS-I beleid van het ministerie, door de DGIS-leiding bepaald dat het toezicht op de kwaliteit van programma-evaluaties zou gelden voor alle 'grote' subsidieontvangers, die samen het leeuwendeel van het totale MFS-I budget van € 2,1 miljard toegewezen hadden gekregen.
Verder werd vastgesteld dat de evaluatie-inspanning van de betreffende organisaties een minimale omvang diende te hebben, opdat het bedrag aan geëvalueerde uitgaven in een redelijke verhouding zou staan tot het ontvangen subsidiebedrag

2.2 Doelstelling

Tegen de hiervoor geschetste achtergrond ligt het voor de hand dat dit meta-onderzoek in de eerste plaats een verantwoordingsfunctie heeft.

Uiteraard is het de bedoeling van de Inspectie dat de bevindingen en conclusies van deze studie daarnaast een bijdrage leveren aan de kwaliteitsverbetering van toekomstig evaluatieonderzoek. In de mate waarin dat het geval is, krijgt deze evaluatie dan naast haar primaire verantwoordingsfunctie tevens een leerfunctie. De mate waarin die tweede functie daadwerkelijk uit de verf komt, hangt grotendeels af van de opstelling van betrokken partijen ten opzichte van de onderzoeksresultaten.

Hoewel verantwoordings- en leerfuncties van evaluatieonderzoek van elkaar kunnen worden *onder*scheiden, is het contraproductief om ze tegenover elkaar te plaatsen alsof ze elkaar zouden hinderen of zelfs uitsluiten. Veeleer is het tegendeel het geval. In de ervaring van de Inspectie laten verantwoorden en leren als gebruiksfuncties van evaluatieonderzoek zich uitstekend combineren. Voor de uitoefening van beide is immers hetzelfde *essentieel*: een gedegen inzicht in werkelijk behaalde resultaten. Dat is niet alleen nodig om verrichte uitgaven te verantwoorden, maar ook bruikbare lessen te kunnen trekken voor toekomstige activiteiten.

Met het primaat voor de verantwoordingsfunctie dat geldt voor deze meta-evaluatie is niet gezegd of gesuggereerd dat andere gebruiksdoelen van evaluatie minder belangrijk zouden zijn. De concentratie op verantwoording vloeit hier eenvoudigweg voort uit het feit dat het analysekader dat IOB dient te hanteren bij de uitvoering van haar mandaat ten opzichte van

het MFS, wordt gevormd door de Regeling Prestatiegegevens en Evaluatieonderzoek Rijksoverheid (RPE).[16] Deze regeling werd uitgewerkt in het kader van de rijksbrede operatie ter verbetering van het resultaatgericht sturen, bekend onder de naam *Van Beleidsbegroting tot Beleidsverantwoording* (*VBTB*).

2.3 Probleemstelling

De probleemstelling brengt onder woorden *wat* men door middel van het onderzoek te weten probeert te komen. De onderzoekdoelstelling geeft weer *waarvoor* men die kennis vervolgens wil gebruiken. Daarmee ligt de onderzoekdoelstelling dus *buiten* het onderzoek zelf, terwijl de probleemstelling er juist de *kern* van vormt. In het onderhavige geval gaat het om de volgende probleemstelling:

In welke mate voldeden programma-evaluaties binnen het Medefinancierings–stelsel aan de eisen van kwalitatief verantwoord evaluatieonderzoek en op welke onderdelen zijn eventueel verbeteringen mogelijk?

Om onderzoekbaar te zijn, moet deze probleemstelling uiteraard worden geoperationaliseerd. Dit gebeurt aan de hand van de criteria voor evaluatiekwaliteit die zijn ontleend aan de RPE, waarmee de drie onderzoekvragen worden hoe:

(i) *valide,*
(ii) *betrouwbaar* en
(iii) *bruikbaar*

de programma-evaluaties van de elf grote MFS-I organisaties zijn die sinds 2007 door IOB werden beoordeeld. Verdere concretisering van de onderzoekvragen vindt plaats in hoofdstuk 4 over de onderzoekaanpak.

2.4 Rapportopzet

Na de inleiding in dit hoofdstuk (2) over de achtergrond en het doel van deze meta-evaluatie is de rest van het rapport als volgt opgezet:

In hoofdstuk 3 wordt het object van deze studie gedefinieerd en afgebakend. Daarbij gaat het om de vragen hoeveel en welke studies van welke subsidieontvangers onder de loep zijn genomen en waarin die studies zich onderscheiden van andere soorten onderzoek die in MFS-kader worden ondernomen.

Daarna gaat hoofdstuk 4 in op het analysekader waarbinnen deze dit onderzoek is uitgevoerd en hoe het onderzoek is aangepakt. Hierbij komen eerst de herkomst en het belang van de gehanteerde beoordelingscriteria aan de orde. Vervolgens wordt uiteengezet op

[16] Ministerie van Financiën. (2001). *Regeling Prestatiegegevens en Evaluatieonderzoek (RPE) Rijksoverheid.* Den Haag.

welke manier de waarderingsscores werden toegekend en hoe de PE-beoordelingsprocedure was ingericht.

Ten slotte geeft hoofdstuk 5 geeft weer tot welke uitkomsten de beoordelingen hebben geleid. Daartoe wordt zowel aandacht besteed aan de omvang van de evaluatie-inspanningen in relatie tot de toegekende subsidiebedragen als aan de methodologische kwaliteiten van de onderzochte programma-evaluaties. Ten slotte komt aan de orde of er sprake is geweest van 'lerend vermogen' bij MFS-organisaties en of dit zichtbaar te maken is.

In bijlagen zijn per MFS-organisatie de voornaamste algemene kenmerken en de scores op de twintig beoordeelde onderdelen van alle 72 programma-evaluaties opgenomen.

De belangrijkste uitkomsten van het onderzoek zijn aan het begin van dit rapport, in hoofdstuk 1, samengevat onder hoofdbevindingen.

3

Onderzoekobject

Het toezicht van de Inspectie op de kwaliteit van evaluatieonderzoek in het medefinancieringsstelsel betrof niet *al* het evaluatieve onderzoek van *alle* MFS-I organisaties, maar een deel ervan. Welk deel dat was komt in dit hoofdstuk aan de orde.

3.1 MFS-I organisaties

Het onderzoekobject van deze meta-evaluatie omvat alle programma-evaluaties van de 'grote' subsidieontvangers onder het MFS-I. Het onderscheid groot–klein werd geïntroduceerd door de Adviescommissie Medefinancieringsstelsel[17] die in 2006 werd ingesteld om de minister van advies te dienen over de toewijzing van de MFS-I middelen die voor de periode 2007 $^{t}/_{m}$ 2010 beschikbaar waren. De cesuur voor het onderscheid tussen kleine en grote aanvragers werd gelegd bij € 2,5 miljoen (gemiddeld) per jaar. Aan organisaties die een aanvraag indienden voor dit of een hoger bedrag werden, in vergelijking met de aanvragers van lagere bedragen, op het gebied van de bedrijfsvoering extra eisen gesteld. Deze eisen maakten deel uit van zogenaamde Organisatietoets of O-toets waaraan alle aanvragers werden onderworpen, en hadden betrekking op de onderdelen 'track record', ondersteuning van partnerorganisaties, innovatiegericht leren, lerend vermogen, doelmatigheid en maatschappelijke oriëntatie.[18]

Toen de minister voor Ontwikkelingssamenwerking op 22 september 2006 bekendmaakte[19] dat aan 58 organisaties een MFS-I subsidie was toegekend, bevonden zich daaronder de volgende elf grote subsidieontvangers:

Tabel 3–1 Subsidietoekenning grote MFS-I organisaties

organisatie	toegekende subsidie (€)
Cordaid	421.830.126
Hivos	260.890.000
ICCO	525.000.000
Oxfam Novib	509.000.000
Terre des Hommes	32.000.000
ETC	12.987.552
International Child Support	23.579.354
IKV/Pax Christi	15.090.000
IUCN	29.895.554
Solidaridad	20.130.000
Woord en Daad	20.190.485
totaal	**1.870.593.071**

17 In de wandeling naar haar voorzitter aangeduid als de Commissie Bikker.

18 *Kwaliteit in ontwikkeling*, eindrapport van de Adviescommissie medefinancieringsstelsel, Bijlage 4 *Aanvraagstramien*, pp. 8-14, september 2006.

19 Brief van de minister voor OS d.d. 22 september 2006, TK: 2006–2007, 27 433, nr. 59.

Van het totale MFS-I budget van € 2.110 miljoen absorbeerden deze elf organisaties samen 89%. De vijf eerstgenoemde waren afkomstig uit het MFP-breed, terwijl de meeste andere eerder subsidie uit het TMF hadden ontvangen.

Tabel 3–1 laat zien dat, naast de sterke concentratie van de middelenverdeling op stelselniveau, ook binnen de categorie 'grote' ontvangers de distributie van subsidies verre van gelijkmatig was. Het gezelschap van elf was dan ook aanmerkelijk minder homogeen dan het gemeenschappelijke etiket 'grote aanvragers' suggereerde. Een aantal van de organisaties[20] waaraan een subsidie tussen € 10 en 30 miljoen was toegekend, wees DSI in juli 2007 op de 'grofmazigheid' van de gehanteerde tweedeling en maakte er bezwaar tegen aan dezelfde evaluatieverplichtingen te worden onderworpen als de (vier) echt grote subsidieontvangers. De briefschrijvers stelden dan ook overleg voor over de nadere invulling van de 'evaluatieverplichting'. Voor zover valt na te gaan is hierop door het ministerie nooit formeel gereageerd.

3.2 Soorten evaluaties

Het toezicht waarmee IOB werd belast, vloeide voort uit de invoering in 2001 van de RPE en had betrekking op de kwaliteit van wat in die regeling wordt aangeduid als *evaluatieonderzoek ex post*. Volgens definitiebepaling 2.2 van de RPE is dat:

Systematisch onderzoek naar de effecten van bestaand beleid, de wijze waarop beleid wordt uitgevoerd en/of de kosten en kwaliteit van geleverde producten en diensten.

Om na te gaan welke binnen het MFS-I (en daaraan voorafgaand het MFP-breed) ondernomen studies aan deze opzet zouden kunnen beantwoorden is het nuttig bij evaluaties onderscheid te maken tussen drie aggregatieniveaus:
- Projectevaluaties
- Individuele programma-evaluaties
- Gezamenlijke programma-evaluaties.

3.2.1 Projectevaluaties

Onder een *projectevaluatie* wordt doorgaans een kleinschalig evaluatieonderzoek verstaan dat zich beperkt tot een individueel project, een afzonderlijke financieringsovereenkomst en/of de relatie met één partnerorganisatie (PO). Voor de beantwoording van de vraag of dergelijke evaluaties aan de kwaliteitsbeoordeling van IOB dienden te worden onderworpen bleken verschillende overwegingen relevant.

Hoewel dit type evaluatie doorgaans plaatsvindt op initiatief van de Nederlandse (donor) NGO – soms *ad hoc* als daar een bijzondere aanleiding toe bestaat, maar meestal op grond van afspraken die vastliggen in de afgesloten financieringsovereenkomst – worden projectevaluaties begeleid en vaak ook gesuperviseerd door de PO waarvan de activiteiten

[20] IKV/Pax Christi, Woord en Daad, IUCN, Solidaridad, ETC en Free Voice: Brief aan DSI d.d. 23 juli 2007.

het object van onderzoek vormen. In sommige gevallen draagt de PO zelfs formeel de eindverantwoordelijkheid voor de kwaliteit van de betreffende evaluaties, die in de regel worden uitgevoerd door externe, maar lokaal gerekruteerde, onderzoekers. Hoe dit ook zij, het ligt voor de hand dat de meeste en meest gedetailleerde informatie over opzet en uitvoering van projectevaluaties zich niet bij de Nederlandse MFS-I organisatie bevindt, maar bij haar partners in de ontwikkelingslanden. Dit is van belang omdat het bij dit kwaliteitsonderzoek niet alleen gaat om een beoordeling van het evaluatierapport, maar om het *hele* onderzoek dat eraan ten grondslag heeft gelegen. Gezien de vaak dominante invloed die het evaluatieontwerp en de sturing van het evaluatieproces op de uiteindelijke onderzoekskwaliteit hebben, is het wenselijk aan die onderdelen bijzondere aandacht te besteden. Dat vergt doorgaans toegang tot de relevante actoren, zoals programmabeheerders en personeel van evaluatie-eenheden. Als die zich bij een partnerorganisatie in een ontwikkelingsland bevinden, is de uitoefening van toezicht vanuit Nederland praktisch gezien niet uitvoerbaar.

Een tweede overweging vormt de omstandigheid dat projectevaluaties volgens de meeste MFO's vóór alles, om niet te zeggen uitsluitend, een *leerfunctie* en geen *verantwoordingsfunctie* vervullen. Vooral tegenstanders van de verantwoordingsfunctie van evaluatieonderzoek neigen er nog wel eens toe het onderscheid tussen beide functies kunstmatig zwaar aan te zetten, alsof zij elkaar per definitie zouden uitsluiten. Hoewel, zoals in het vorige hoofdstuk al werd betoogd, dit geenszins het geval is, moet wel worden onderkend dat de leer- en verantwoordingsfuncties zich meestal op verschillende doelgroepen richten. Waar het afleggen van verantwoording vooral wordt geassocieerd met categorieën van belanghebbenden als opdrachtgevers en donateurs, bestaat de primaire doelgroep van de leerfunctie doorgaans uit vormgevers, uitvoerders en begunstigden van de geëvalueerde activiteiten. Een belangrijk gevolg van dit verschil is dat zogenaamde *leerevaluaties* kunnen volstaan met een veel beperktere hoeveelheid feitelijke en achtergrondinformatie over evaluatieobject en -context, omdat de beoogde gebruikersgroep van de onderzoekrapportage bestaat uit ingewijden. Voor buitenstaanders zijn dergelijke onderzoeksrapporten veelal moeilijk of zelfs niet toegankelijk, omdat ze bij de lezer te veel bekend veronderstellen. Verder zijn 'leerevaluaties' doorgaans sterk op de toekomst gericht, omdat niet zozeer het al dan niet voortbestaan van het evaluatieobject ter discussie staat, maar de *verbetering* ervan. Dit brengt met zich mee dat de nadruk (bedoeld of onbedoeld) vaak sterker komt te liggen op de formulering van nieuw beleid dan op de empirische vaststelling van in het verleden bereikte resultaten van bestaand beleid.

De toekomstgerichtheid wordt in veel gevallen verder versterkt doordat deze projectevaluaties voor de betrokkenen een belangrijke rol spelen bij de voorbereiding van managementbeslissingen over de verlenging of beëindiging van contracten, relaties met partnerorganisaties, personele inzetten en dergelijke. Deze beheersfunctie van projectevaluaties bevordert welhaast onvermijdelijk de al aanwezige toekomstgerichtheid van de betreffende onderzoeken en rapportages ten koste van het empirische karakter.

3.2.2 Programma-evaluaties

Programma-evaluaties verschillen van projectevaluaties doordat ze een bredere opzet hebben en een omvangrijker onderzoekobject bestrijken dan het individuele project of de relatie met één enkele PO. Ze zijn te onderscheiden *in individuele* en *gezamenlijke* PE's. Bij een individuele PE gaat het om één afzonderlijke MFS-I organisatie die een reeks van haar eigen activiteiten evalueert of laat evalueren. Gezamenlijke PE's worden door meerdere organisaties ondernomen in het kader van hun samenwerking binnen het besloten gedeelte van het *Platform Kwaliteitshuis*.[21] Dit vormt onderdeel van Partos, de in 2004 opgerichte brancheverenigingen voor Nederlandse particuliere organisaties die actief zijn op het terrein van de internationale samenwerking. Het Kwaliteitshuis omvat een reeks van gezamenlijke activiteiten die zich richten op kwaliteitsverbetering van het werk van Partosleden, waaronder programma-evaluaties.

Tijdens de subsidieperiode 2003–2006 was al gebleken dat er geen algemeen aanvaarde definitie van het begrip programma-evaluatie bestond, waardoor het dan ook niet voor iedereen hetzelfde inhield. Aan de hand van een aantal aangeleverde studies zijn toen geleidelijk de contouren van het PE-begrip aangescherpt totdat een omschrijving en afbakening tot stand kwam die in de praktijk voor alle betrokken partijen hanteerbaar was. Hierbij bleken, behalve uiteraard de toetsbaarheid aan de kwaliteitscriteria uit de RPE (zie hoofdstuk 4), vooral drie dimensies van belang:

- De term **programma** blijkt in de praktijk te fungeren als een containerbegrip voor uiteenlopende soorten verzamelingen. Een programma-evaluatie dient dan ook te worden beschouwd als een onderzoek waarvan het object het niveau van het individuele ontwikkelingsproject, de individuele financieringsovereenkomst en/of de relatie met een individuele partner overstijgt. Dit houdt in dat een PE-onderzoek in ieder geval betrekking moet hebben op een *reeks* van activiteiten die een of meer van de volgende kenmerken gemeen hebben:
  - sector (bijvoorbeeld onderwijs);
  - thema (bijvoorbeeld milieu);
  - land (bijvoorbeeld Burkina Faso);
  - regio binnen een land (bijvoorbeeld Noord-Irak);
  - landen binnen een regio (bijvoorbeeld Centraal-Azië);
  - interventiestrategie (bijvoorbeeld maatschappijontwikkeling),

  waarbij het er niet toe doet of de betreffende MFS-I organisaties de geëvalueerde reeks activiteiten buiten het kader van het evaluatieonderzoek ook een *programma* noemen.

- Een PE hoort gericht te zijn op **resultaatvaststelling**. Die resultaten kunnen op *output*, *outcome* of *impact*niveau liggen, maar zijn in ieder geval duidelijk te onderscheiden van allerlei procesvariabelen zoals intern-organisatorische maatregelen, klantentevredenheidsscores, (back)donorrelaties, enz. Gezien vanuit de RPE die de basis vormt voor de IOB-kwaliteitsbeoordeling, bevinden dergelijke 'procesresultaten' zich vooral aan de *input*-kant van de interventielogica.

[21] Leden van het besloten gedeelte van het Platform Kwaliteitshuis zijn de organisaties Cordaid, Hivos, ICCO, Oxfam Novib en Plan Nederland. De laatstgenoemde organisatie ontving onder het MFS-I geen subsidie.

De vaststelling van impact, in de betekenis van langere-termijn resultaten, blijkt in de meeste gevallen te hoog gegrepen te zijn. Aan de andere kant wordt bij evaluatieonderzoek dat zich beperkt tot de meting van output, opgevat als de directe (fysieke) prestatie van de NGO, de lat in het algemeen te laag gelegd. Zowel nuttig (voor verantwoordingsdoeleinden) als (in de meeste gevallen) realistisch zijn PE's die hun resultaatmeting concentreren op het outcomeniveau, dat wil zeggen op de maatschappelijke effecten die (mede) dankzij de geleverde (output)prestaties zijn bereikt.

- Het toezichtmandaat van IOB betreft de besteding van **MF-fondsen** uit de OS-begroting. Dit betekent dat, om voor een beoordeling in aanmerking te komen, de in een PE onderzochte activiteiten ten minste gedeeltelijk met *medefinancieringsmiddelen* dienen te zijn gefinancierd. Subsidies uit andere budgetten van het DGIS, zoals noodhulp, maar ook bijdragen van andere donoren blijven daarom buiten beschouwing. De omvang van het beslag op de MF-middelen is een voor de beoordeling relevant kenmerk met het oog op de vaststelling van de 'dekkingsgraad' die in § 3.3 aan de orde komt.

In het kader van deze meta-evaluatie zijn in totaal 72 PE's beoordeeld, waaronder 3 gezamenlijke en 69 individuele. Deze waren als volgt over de verschillende organisaties verdeeld.

Tabel 3–2 Aantallen Programma-evaluaties

organisatie	aantal PE's
Cordaid	11
Hivos	7
ICCO	9
Oxfam Novib	12
Terre des Hommes	4
ETC	2
International Child Support	8
IKV/Pax Christi	9
IUCN	1
Solidaridad	3
Woord en Daad	3
subtotaal individuele PE's	69
gezamenlijke PE's (Partos)	3
Totaal	**72**

Ten slotte dient vanuit het perspectief van het gebruik van de uitkomsten van evaluatieonderzoek nog een belangrijk onderscheid te worden gemaakt. Waar de resultaten van projectevaluaties primair zo niet uitsluitend hun toepassing vinden op het niveau van individuele interventies, worden de uitkomsten van programma-evaluaties vooral terugge-

koppeld naar een *beleidsniveau* dat het individuele project overstijgt. Op dat niveau worden de onderzoekuitkomsten, als het goed is, benut voor nadere invulling, bijsturing of zelfs complete herijking van het gevoerde sectorale, thematische, regionale of strategische *beleid*. Hoewel in deze meta-evaluatie is getracht een oordeel te vormen over de *bruikbaarheid* van de voorgelegde PE's, viel de mate waarin de beoordeelde studies daadwerkelijk door betrokkenen zijn *gebruikt*, buiten het bestek van het onderzoek.

Individuele PE's

Vergeleken met de MFP-breed periode (2003–2006) is onder het MFS-I het aantal individuele PE's in verhouding tot de financiële programma/stelselomvang aanzienlijk toegenomen. Onder het MFP-breed waren dat er bij een programmaomvang van ca. € 1,6 miljard nog 22. Tabel 3–2 laat zien dat in de periode 2007–2011, toen de stelselomvang bijna € 1,9 miljard bedroeg, 69 individuele programma-evaluaties werden uitgevoerd en op hun kwaliteiten beoordeeld.

Ondanks de afbakening van het PE-begrip vertonen deze onderzoeken nog een vrij grote verscheidenheid in omvang, opzet en uitvoering, zoals te zien valt in bijlage 2, waarin de belangrijkste kenmerken van de bestudeerde onderzoeken zijn weergegeven.

Gezamenlijke PE's

Met de beëindiging van het mandaat van de onafhankelijke Stuurgroep Evaluatie Medefinancieringsprogramma in 2002 werden de MFO's bij de invoering van het MFP-breed en de start van de nieuwe subsidieperiode 2003 – 2006 zelf geheel verantwoordelijk voor de opzet en uitvoering van resultaatgericht evaluatieonderzoek.

Daartoe stelde een speciale Werkgroep Evaluatie & Beleid (E&B) van het GOM eind 2002 een gezamenlijke meerjarige onderzoekagenda op. Deze gold voor de hele subsidieperiode en bevatte zeven onderwerpen voor evaluatieonderzoek, waaraan de GOM-partners in wisselende samenstellingen (variërend van twee tot alle vijf MFO's) zouden deelnemen. Kernvraag van iedere gezamenlijke programma-evaluatie zou worden: *Wat zijn de resultaten die bereikt zijn (op het niveau van bereik en outcome op doelgroepniveau) en hoe zijn deze te beoordelen?*[22] Voor elk onderzoek werd een Coördinatiegroep geformeerd met vertegenwoordigers van de deelnemende MFO's en één van de GOM-directeuren als voorzitter.
Van de zeven geplande onderzoeken zijn er uiteindelijk vijf uitgevoerd en twee geschrapt. De eerste die wel werd uitgevoerd (*Basiseducatie*) resulteerde na een moeizaam verlopen evaluatieproces in een eindrapport dat volgens de betrokken MFO's (Plan en Novib) de toets der (kwaliteits)kritiek niet kon doorstaan en daarom werd bestempeld tot "intern leerdocument".

[22] GOM (2003): *Medefinancieringsprogramma (MFP)-evaluatie 2003–2006*, p. 5. In de latere actualisering van de onderzoeksagenda werd deze kernvraag aangescherpt tot: *Welke resultaten zijn door de interventies van de MFO's en hun partner-organisaties bereikt (op verschillende niveaus) en hoe zijn deze resultaten te beoordelen in termen van relevantie en effectiviteit?* (MBN (2004): *Plan van aanpak en kwaliteitscriteria tweede fase MFP-evaluaties 2003–2006: november 2004 – juni 2006*.

Dit laatste gebeurde mede op advies van een inmiddels ingestelde *Evaluatie Referentie Groep* (ERG)[23] bestaande uit een zevental onafhankelijke, externe deskundigen met als mandaat de beoordeling van de kwaliteit van proces en resultaten van de gezamenlijke GOM en latere MBN-evaluatiestudies. Hiertoe richtte men zich op vier gebieden: de overkoepelende *Terms of Reference* (ToR), het evaluatieontwerp, het eindproduct van de evaluatie en het evaluatieproces. De ERG begeleidde en becommentarieerde na het eerste onderzoek eveneens de vier nog resterende MBN-studies die wel zijn uitgemond in volwaardige PE-rapporten. Uit de IOB meta-evaluatie van 2006 bleek dat deze onafhankelijke referentiegroep een heilzaam effect heeft gehad op de methodologische kwaliteit van de gezamenlijke PE's.

Met de start van het Medefinancieringsstelsel in 2007 werd het MBN opgeheven en brachten de leden van het netwerk hun gezamenlijke evaluatieprogrammering, die zij beschouwden als hét instrument voor het samen afleggen van verantwoording over bereikte resultaten[24], onder in het Kwaliteitshuis van Partos. De planning voor de periode 2007–2010 omvatte een vijftal studies, waarvan de laatste (*Geweld tegen Vrouwen*) uiteindelijk niet doorging omdat er bij het aflopen van de subsidieperiode nog geen voor alle betrokken partijen aanvaardbare onderzoekopzet was ontwikkeld. De op één na laatste (*Capaciteitsopbouw*) werd wel uitgevoerd, maar uit de IOB-beoordelingsprocedure teruggetrokken omdat de studie essentiële elementen miste die onontbeerlijk zijn voor een beoordeling als programma-evaluatie. Van de drie wel voltooide en beoordeelde gezamenlijke PE's – deze betroffen respectievelijk *Conflicttransformatie*, *Microfinanciering* en *Inheemse Volken* – zijn de belangrijkste kenmerken opgenomen in bijlage 2.

3.3 'Evaluatiedekking'

De hiervoor weergegeven uiteenzetting over de evaluatiepraktijk onder het MFS-I resulteert in de vaststelling dat het totale onderzoekobject van deze meta-evaluaties bestaat 69 individuele + 3 gezamenlijke = 72 programma-evaluaties.

Anders dan bij het kwaliteitstoezicht op het MFP-breed werden er ten aanzien van de evaluatie-inspanningen van de elf 'grote' subsidieontvangers onder het MFS-I niet alleen *kwalitatieve* maar ook *kwantitatieve* verwachtingen geformuleerd. De bedoeling hiervan was te verzekeren dat de te leveren evaluatieprestaties een minimale omvang zouden hebben. Daartoe werd het bedrag aan MFS-I middelen dat met PE-onderzoek bestreken diende te worden, uitgedrukt in een percentage van de toegekende subsidie voor de hele periode 2007 – 2010.

Voor de vaststelling van het minimale 'dekkingspercentage' werd aansluiting gezocht bij de al genoemde evaluatierichtlijnen van de RPE. Deze uit 2001 daterende regeling was na een evaluatie begin 2006 herzien en op een aantal punten versoepeld, waaronder de frequentie

[23] Soms wordt deze afkorting uitgelegd als staande voor ***Externe*** *Referentie Groep*. Beide betekenissen sporen overigens met de werkelijkheid: de Referentiegroep was zowel *extern* ten opzichte van het MBN (en later Partos) als ingesteld ten behoeve van *evaluaties*.

[24] Jaarverslag Partos 2009, p.11

waarmee evaluatieonderzoek dient te worden ondernomen. Gold hiervoor in de oorspronkelijke Regeling nog een standaardfrequentie van eenmaal in de vijf jaar, in de herziene versie van 2006 werd deze strakke periodiciteit geflexibiliseerd. In plaats van één standaardfrequentie werd 'een zekere regelmaat – bijvoorbeeld eenmaal in de vier, vijf of zeven jaar' voorgeschreven[25]. Als men de laagste van deze frequenties als uitgangspunt neemt, dient per jaar 15% van een beleid te worden geëvalueerd om het in zeven jaar helemaal, dat wil zeggen voor 100% (105 om precies te zijn), 'af te dekken', zoals dat in de RPE heette. Rekening houdend met de heterogeniteit van de groep van elf subsidieontvangers, waarbij programma-evaluaties voor sommige organisaties gesneden koek waren terwijl ze voor andere een nieuw en onbekend verschijnsel vormden, lag het voor de hand in plaats van één uniform percentage voor iedereen, een bandbreedte te hanteren. Deze werd bepaald op 10 – 15% per jaar. Voor de hele subsidieperiode van vier jaar betekende dit dus dat het bedrag aan geëvalueerde MF-uitgaven 40 – 60% zou vormen van de in 2006 toegekende MFS-I subsidie.

Als de te evalueren uitgaven *alleen* gefinancierd zouden mogen zijn met uit de MFS-I subsidie afkomstige middelen, zouden de organisaties vóór de afloop van de subsidieperiode (eind 2010) nooit 40 – 60% ervan op maatschappelijk effecten (outcomeniveau) hebben kunnen evalueren. Daarvoor heeft een te groot deel van de activiteiten van de MFS-I organisaties een te lange doorlooptijd: de activiteitencyclus – identificeren → formuleren → beoordelen → committeren → contracteren → uitvoeren → evalueren →(identificeren) – is op veel werkterreinen gewoonweg te lang om binnen een periode van vier jaar volledig en zorgvuldig te kunnen worden afgewikkeld. Dat dit verschijnsel ook in de RPE als een mogelijke complicatie wordt onderkend valt op te maken uit de bepaling in het hoofdstuk over de afdekking van beleid, dat "in het algemeen een beoordeling achteraf van de doeltreffendheid en doelmatigheid van beleid pas kan worden opgemaakt op het moment dat de resultaten van het betreffende beleid redelijkerwijs verwacht mogen worden. Eerder is het uitvoeren van een evaluatieonderzoek ex post niet zinvol."[26]
Naast het probleem van de doorlooptijd was er het gegeven dat veruit de meeste organisaties al uit de voorlopers van het MFS-I (MFP-breed en TMF) subsidie hadden ontvangen, waarvan zij bij de start van het MFS-I op 1 januari 2007 een deel nog niet hadden uitgegeven. Sommige organisaties bleken er tot een jaar over te doen om de eerste tranche van de hun toegekende nieuwe MFS-I subsidie op te vragen.

Het pragmatische antwoord op deze complicaties werd de afspraak dat voor de 'evaluatiedekking' ook onderzoek van activiteiten ontplooid vóór 2007 in aanmerking kon worden gebracht mits (i) het ging om de evaluatie van *doorlopende* activiteiten waarvan de financiering onder het MFS-I gewoon werd voortgezet en (ii) de eerdere financiering eveneens uit MF-subsidies had plaatsgevonden (MFP, MFP-breed of TMF). De uitbreiding tot langere evaluatieperioden met een begindatum vóór 1 januari 2007 bleek niet overbodig: van de 72 beoordeelde PE's betrof slechts ongeveer een kwart (19) uitsluitend activiteiten die binnen

[25] Nota van toelichting op de RPE 2006 (pp. 11-12), bijlage bij brief van de minister van Financiën aan de Tweede Kamer d.d. 18 april 2006, TK: 2005 – 2006, 29 950, nr. 40.

[26] RPE 2001, p. 29.

de MFS-I periode waren ontplooid. De evaluatieperioden van alle overige onderzoeken gingen verder terug in de tijd, tot een maximum van twaalf jaar (1995).

Het gevolg van een en ander is dat het begrip 'evaluatiedekking' hier niet beperkt wordt tot de verantwoording van bestedingen van MFS-I subsidies in strikte zin. Evaluatiedekking' is in deze meta-evaluatie op te vatten als de *verhouding* tussen

- enerzijds de totale bedragen aan MF-uitgaven – in ruime zin[27] – die de elf 'grote' subsidieontvangers door middel van programma-evaluaties hebben verantwoord, en
- anderzijds de omvang van de onder het MFS-I aan diezelfde organisaties toegekende subsidiebedragen.

Met deze invulling van het begrip 'evaluatiedekking' is niet meer beoogd dan te bevorderen dat de 'grote' subsidieontvangers onder de MFS-organisaties tegenover het toegekende subsidiebedrag een in kwantitatief opzicht serieuze *evaluatie-inspanning* zouden ontplooien.

3.4 Verantwoordingsvormen

Hoewel PE's belangrijke instrumenten zijn om verantwoording af te leggen over het gevoerde beleid, de uitgaven die ervoor zijn verricht en de resultaten die ermee zijn behaald, vormen ze niet de enige manier waarop MFS-organisaties rekenschap afleggen.

Alle hier besproken subsidieontvangers produceren ook jaarverslagen met jaarrekeningen voorzien van accountantsverklaringen, waarmee zij in het bijzonder de rechtmatigheid van hun bestedingen aantonen. Daarnaast worden bij veel organisaties regelmatig onderzoeken naar klantentevredenheid uitgevoerd, waarmee onder andere de kwaliteit van externe contacten en communicatie tegen het licht wordt gehouden.

Verder wordt ook niet alleen verantwoording afgelegd tegenover de geldschieters: particuliere en gouvernementele donoren. In veel NGO-relaties is het gebruikelijk dat Nederlandse MFS-organisaties tegenover de rekenschap zij van hun partners ontvangen omgekeerd aan hun PO's verantwoording afleggen over gevoerd beleid, geleverde bijdragen en het nakomen van afspraken.

En ten slotte bestaat er naast de traditionele *upward accountability* een groeiende praktijk van zogenoemde *downward accountability*: de niet altijd eenvoudig te organiseren, maar wel door velen noodzakelijke geachte verantwoording van de samenwerkende partners uit Noord én Zuid tegenover de doelgroepen voor wie de gezamenlijke inspanningen uiteindelijk bedoeld zijn.

Deze meta-evaluatie beperkt zich tot de verantwoordingsfunctie van programma-evaluaties. Andere hier genoemde vormen van 'verantwoording' maken dus geen deel uit van het onderzoek.

[27] Met ruime zin wordt gedoeld op het feit dat de verantwoorde uitgaven niet alleen met middelen uit het MFS-I hoeven te zijn gefinancierd, maar ook kunnen zijn betaald met subsidies uit eerdere medefinancieringsprogramma's, zoals Het MFP, MFP-breed of de TMF.

4

Onderzoekaanpak

De beoordeling van de 72 individuele en gezamenlijke programma-evaluaties vond plaats aan de hand van een aantal criteria en in verschillende stappen. Dit hoofdstuk beschrijft wat de gebruikte beoordelingscriteria inhouden en waarom ze relevant zijn. Daarna wordt ingegaan op de gehanteerde waarderingsmethode en op het verloop van de beoordelingsprocedure.

4.1 Regeling Prestatiegegevens en Evaluatieonderzoek (RPE)

Zoals eerder vermeld kwam het beoordelingsstramien dat IOB hanteerde om invulling te geven aan het toezicht op de kwaliteit van programma-evaluaties niet uit de lucht vallen. Het was geënt op de door de minister van Financiën in februari 2001 geïntroduceerde Regeling Prestatiegegevens en Evaluatieonderzoek (RPE) Rijksoverheid. Deze vormt een uitwerking van de Comptabiliteitswet, die bepaalt dat de ministers verantwoordelijk zijn voor het beleid dat aan hun begroting ten grondslag ligt.

Mede naar aanleiding van kritische onderzoeken van de Algemene Rekenkamer startte in 1998 de rijksbrede operatie *Van Beleidsbegroting Tot Beleidsverantwoording* (VBTB) met als 'ambitie om het resultaatgerichte sturingsmodel een stevige impuls te geven'. Essentieel daarvoor was de koppeling van financiële en beleidsinformatie. Om die te bevorderen kwamen in de departementale begrotingen de volgende drie vragen centraal te staan:

- Wat willen we bereiken?
- Wat gaan we daarvoor doen?
- Wat mag dat gaan kosten?

Om na afloop de doeltreffendheid en doelmatigheid van het gevoerde beleid te kunnen verantwoorden moeten de *ex post* pendanten van bovenstaande *ex ante* begrotingsvragen worden beantwoord:

- Wat is er bereikt?
- Wat is daarvoor gedaan
- Wat heeft dat gekost?

Hiervoor is gedegen evaluatieonderzoek noodzakelijk. Alleen daarmee kunnen de resultaten van het gevoerde beleid met voldoende zekerheid worden vastgesteld om de basis te leveren voor besluitvorming over voortzetting, bijsturing of stopzetting. Op die manier wordt evaluatieonderzoek binnen de beleidscyclus tevens een instrument voor sturing. en wordt de evaluatiekwaliteit mede bepalend voor de kwaliteit van de besluitvorming over het beleid.[28]

Doel van de RPE die geënt werd op de VBTB-operatie, was te verzekeren dat beleidsevaluatie van uitgaven (zoals die ten behoeve van het MFS) die worden verantwoord via departementale begrotingen, zou voldoen aan elementaire methodische kwaliteitseisen. De Interdepartementale Begeleidingsgroep Prestatiegegevens en Beleidsevaluatie (IBP) die de

[28] RPE 2001, pp. 10, 33.

Regeling opstelde, bedacht deze eisen niet zelf, maar ontleende ze aan relevante literatuur over methoden en technieken van sociaalwetenschappelijk onderzoek.[29]

4.2 Beoordelingscriteria

De gangbare maatstaven voor toegepast sociaalwetenschappelijk onderzoek, waartoe beleidsevaluatie algemeen wordt gerekend, zijn:

Validiteit

Deze heeft betrekking op de algemene geldigheid van het onderzoekontwerp, inclusief de gehanteerde begrippen en instrumenten. De centrale vraag is of men met het onderzoek daadwerkelijk meet wat men denkt en beoogt te meten? Hiervoor is het noodzakelijk *systematische fouten* te vermijden, vooral bij de operationalisering van abstracte kenmerken van resultaten. Het ligt dan ook voor de hand dat er een positief verband bestaat tussen validiteit en de methodologische kwaliteit van het *onderzoekontwerp*. Dat ontwerp vindt zijn neerslag vindt in de Terms of Reference waarvoor de verantwoordelijkheid uiteraard geheel berust bij de opdrachtgever van de evaluatie. Een voorwaarde voor validiteit is verder een coherent en transparant verband tussen gegevens, analyse, conclusies en aanbevelingen.

Betrouwbaarheid

Deze eigenschap heeft vooral betrekking op de stabiliteit van de onderzoekuitkomsten, dat wil zeggen op de vraag of herhaling van het onderzoek hetzelfde resultaat zou opleveren. De betrouwbaarheid wordt bepaald door de mate waarin het lukt *toevalsfouten* bij de uitvoering van het onderzoek te vermijden. Essentieel hiervoor is de representativiteit van de onderzochte eenheden terwijl verder de onafhankelijkheid van zowel de onderzoekers als het gebruikte onderzoekmateriaal een belangrijke factor vormt. Waar de ontwerpfase grotendeels bepalend is voor de validiteit, hangt de betrouwbaarheid van evaluatieonderzoek het sterkst af van de kwaliteit van de uitvoeringsfase en dus van de ingeschakelde onderzoekers.

Bruikbaarheid

Dit derde kenmerk is van belang omdat beleidsevaluatie doorgaans geen fundamenteel maar toegepast onderzoek is, dat wil zeggen dat het een buiten het onderzoek zelf gelegen (extern) doel dient zoals verantwoording afleggen of lessen trekken. De bruikbaarheid van evaluatieonderzoek hangt allereerst af van de mate waarin het aan de eerste twee kwaliteits criteria, validiteit en betrouwbaarheid, voldoet. Daarnaast is voor opdrachtgevers en gebruikers van belang of het onderzoek hun specifieke informatiebehoeften bevredigt en of de implementatie van de adviezen die op het onderzoek zijn gebaseerd, binnen hun vermogen ligt. Verder dient bruikbaar ex post evaluatieonderzoek zoals opgevat binnen de RPE, in ieder geval beoordeling op doeltreffendheid en doelmatigheid mogelijk te maken.

[29] Een van de belangrijkste auteurs wier werk daartoe werd geraadpleegd was de voormalige hoogleraar methodenleer P.G. Swanborn.

Hoe relevant validiteit, betrouwbaarheid en bruikbaarheid ook zijn als algemene criteria voor de beoordeling van evaluatieonderzoek, voor directe toepassing als maatstaven ter bepaling van methodologische kwaliteiten zijn ze te grof en te multidimensionaal. Ze moeten dus nader worden geoperationaliseerd en gespecificeerd tot een hanteerbaar toetsingsinstrument. De ontwikkeling hiervan was al gestart in het kader van eerdere IOB-onderzoeken.[30] Het daarbij ontwikkelde stramien ligt ten grondslag aan de zogenaamde *Beoordelingslijst* die is toegepast in de onderhavige studie, zij het dat in de loop van de tijd en op basis van voortschrijdend inzicht enkele voornamelijk redactionele aanpassingen en verduidelijkingen werden aangebracht. Daarbij bleef echter de basisstructuur, die men zich kan voorstellen als een piramide zoals weergegeven in Figuur 4–1, onveranderd.

Figuur 4–1 *Beoordelingspiramide*

Op grond van ervaringen die gedurende de vier jaar van de MFP-breed subsidieperiode met de kwaliteitsbeoordeling van programma-evaluaties was opgedaan, werd eind 2006, na overleg met de evaluatievertegenwoordigers van de toenmalige (vijf) MFO's besloten de gehanteerde Beoordelingslijst te stroomlijnen. Hierbij werd het aantal kwaliteitsindicatoren teruggebracht van 14 naar 10 en het aantal beoordelingscomponenten van 32 naar 20, waarmee de overzichtelijkheid en hanteerbaarheid toenamen. De actuele, herziene versie van de lijst, inclusief beknopte toelichtingen op elk van de 20 componenten zoals die gedurende de hele subsidieperiode van het MFS-I werd gehanteerd, is opgenomen in bijlage 3. Om vergelijking met de beoordelingsresultaten uit de MFP-breed periode mogelijk te maken, moesten deze worden omgezet naar de systematiek van de herziene Beoordelingslijst. Hoe deze conversie van de oude naar de nieuwe lijst heeft plaatsgevonden, wordt verantwoord in bijlage 4.

De treden van de piramide verbeelden de verschillende beoordelingsniveaus, waarbij van boven naar beneden elke laag is te beschouwen als een verdere *operationalisering* van de eraan voorafgaande. Zo werden de drie RPE-kwaliteitscriteria validiteit, betrouwbaarheid en bruikbaarheid geoperationaliseerd door middel van tien *indicatoren* en zijn deze op hun beurt nader geconcretiseerd in 20 componenten. De bedoeling van deze steeds verdergaan-

[30] IOB (2004): *Onderzoek naar de kwaliteit van in 2002 afgeronde decentrale evaluaties*. Den Haag: IOB Evaluaties nr. 298 en IOB (2005): *Onderzoek naar de kwaliteit van in 2003 afgeronde decentrale evaluaties*. Den Haag: IOB Werkdocument.

de detaillering is geweest om de te hanteren indicaties voor evaluatiekwaliteit zo *SMART*[31] mogelijk te krijgen. Dat wil dus zeggen: steeds concreter en eenduidiger. Het proces van steeds verdergaande concretisering kan overigens niet eindeloos worden voortgezet. Dat komt doordat, gaande van boven naar beneden, de variabelen niet alleen steeds gedetailleerder worden, maar ook steeds talrijker: één enkele variabele volstaat namelijk zelden om een eraan voorafgaand fenomeen voldoende nauwkeurig en betrouwbaar 'zichtbaar' te maken. En naarmate de specificiteit van een variabele toeneemt, neemt de toepasbaarheid in verschillende soorten onderzoekssituaties en voor uiteenlopende evaluatie-objecten af. Kortom, met de weergegeven vier niveaus bleek in praktische zin de grens van de mogelijke detaillering wel bereikt.

4.2.1 Validiteit

Voor de operationalisering van het criterium validiteit zijn vier indicatoren geïdentificeerd, elk bestaande uit twee componenten.

Probleemstelling

Aangezien de meeste programma-evaluaties een groot domein met veel (verschillende) activiteiten bestrijken, worden probleemstellingen veelal op een betrekkelijk hoog abstractie- en/of aggregatieniveau geformuleerd. Om te kunnen worden beantwoord moet de probleemstelling worden uitgewerkt in *onderzoekvragen*, waarbij het zaak is deze zo concreet en eenduidig mogelijk te formuleren. Als dat niet gebeurt, blijft namelijk vaak onduidelijk of ze met de uitkomsten van het onderzoek wel of niet (adequaat) worden beantwoord. Verder is het van belang bij de opstelling van de onderzoekvragen hoofd- en bijzaken scherp te blijven onderscheiden. Daarbij dient de uiteindelijke opsomming te worden beperkt tot de (hoofd)vragen die door het voorgenomen onderzoek, gegeven de beschikbare tijd en middelen, daadwerkelijk kunnen worden beantwoord.

Minstens even belangrijk is duidelijkheid over de maatstaven die worden aangelegd om tot de beoogde oordelen te komen. Daartoe moeten de *evaluatiecriteria* niet alleen expliciet worden benoemd maar ook ondubbelzinnig worden gedefinieerd. Het feit dat gangbare *termen* als relevantie, doeltreffendheid en doelmatigheid in brede kring zijn ingeburgerd, betekent niet dat de *begrippen* waarnaar ze verwijzen ook eenduidig zijn, dat wil zeggen: voor iedereen dezelfde lading hebben. Dit is cruciaal, omdat misverstanden tussen de bij een evaluatie betrokken partijen op dit punt verstrekkende gevolgen kunnen hebben voor het gemeenschappelijk begrip van de onderzoekresultaten, en daarmee voor het gebruik ervan.

[31] SMART is een Engelstalig acroniem dat staat voor: Specifiek, Meetbaar, *Attainable* [Bereikbaar, haalbaar], Relevant en Tijdgebonden. Het somt de eigenschappen op die een ideale indicator in zich zou moeten verenigen.

Evaluatieobject

Bij de omschrijving van het *evaluatieobject* gaat het in een programma-evaluatie doorgaans om de identificatie en afbakening van de operationele populatie, dat wil zeggen de verzameling onderzoekeenheden waarvoor de gevraagde evaluatieve uitspraken zullen (moeten) gelden. Voor een helder begrip van het evaluatieobject zijn *kwantitatieve* gegevens over typen activiteiten, doelgroepen, locaties, perioden en uitgaven, enz. van belang. Daarnaast is het nodig goede *kwalitatieve* informatie te geven over de (beoogde) werking van de geëvalueerde interventie(s), zodat duidelijk is aan welke doelstelling(en) bereikte resultaten dienen te worden getoetst. Zonder heldere omschrijving en afbakening van het evaluatieobject is het voor niet-ingewijden lastig om zich een beeld te vormen van *wat* er nu eigenlijk is onderzocht en wordt de verantwoordingsfunctie van het onderzoek aangetast.

Een goede beschrijving en analyse van de fysieke en institutionele omgeving waarin de geëvalueerde activiteiten zich afspelen, is onontbeerlijk voor een correcte duiding van behaalde resultaten.[32] Immers, veel van de interventies zijn naar hun aard context-gevoelig. In het bijzonder geldt dit wanneer ze door de promotie van emancipatie en rechten van (achtergestelde) doelgroepen gevestigde belangen aantasten. Een *contextanalyse* is dus meer dan een simpele decorschildering. Ze dient de krachten in beeld te brengen die een significante invloed op het resultaat van de interventies (kunnen) hebben gehad.

Beleidstheorie

PE's die in MFS-kader worden ondernomen, behoren te zijn opgezet als theorie-gestuurd evaluatieonderzoek. Immers, de subsidiebesteding moet worden verantwoord tegen de achtergrond van de beleidsvoornemens en doelstellingen waarop de subsidietoekenning was gebaseerd. Elke verantwoordingsevaluatie dient dan ook een uiteenzetting te bevatten over de onderliggende *interventielogica* met de veronderstellingen over causale en finale relaties die aan de onderzochte interventies ten grondslag hebben gelegen. In deze beleidstheorie dienen de doelmiddelenhiërarchie en de verschillende resultaatniveaus duidelijk herkenbaar te zijn.

Om vast te kunnen stellen in hoeverre resultaten daadwerkelijk werden bereikt, zijn doorgaans *indicatoren* nodig, dat wil zeggen variabelen waarvan de waarden de veranderingen die het gevolg zijn van de geëvalueerde activiteiten, zo nauwkeurig mogelijk weerspiegelen. De selectie van goede indicatoren is niet altijd gemakkelijk maar wel altijd van groot belang, zeker wanneer de nagestreefde doelstellingen abstract zijn. Bij de beoordeling van de gebruikte indicatoren gaat het niet alleen om de mate waarin en de consistentie waarmee deze werden toegepast, maar ook in hoeverre zij als specifiek, meetbaar, bereikbaar, relevant en tijdgebonden (*SMART*) kunnen worden aangemerkt.

Analyse

Voor een valide gegevensanalyse moeten de selectie van *gegevensbronnen*, de extractie van gegevens en hun verwerking tot bevindingen op een zorgvuldige en vooral transparante

[32] De evaluatiebenadering die bekend staat als de *Realist Evaluation* (Pawson en Tilly) gaat er zelfs vanuit dat voor de bepaling en verklaring van het eindresultaat de context veelal belangrijker is dan de onderzochte interventie zelf.

manier worden uitgevoerd. De transparantie-eis houdt in dat een geïnteresseerde maar oningewijde lezer aan de hand van de rapportage in staat moet zijn te volgen hoe en waarom specifieke data werden geselecteerd en gecombineerd.

De *conclusies* die op basis van de gepresenteerde analyse worden getrokken, moeten daar logisch uit volgen. Ze dienen dus voldoende te zijn gefundeerd. Wat in dit verband voldoende is, hangt vooral van de aard van de conclusies af. Naarmate deze specifieker (informatiever) zijn, dienen ze uitvoeriger te zijn onderbouwd.

4.2.2 Betrouwbaarheid

Evenals bij het criterium validiteit zijn voor de operationalisering van betrouwbaarheid vier indicatoren, elk gesplitst in twee componenten, gehanteerd.

Onderzoekmethoden

Voor de beoordeling op dit onderdeel wordt in de eerste plaats gekeken naar de identificatie en beschrijving van de toegepaste *onderzoekmethoden*. Vervolgens is het de bedoeling dat door de onderzoekers wordt uitgelegd *waarom* bepaalde methoden en technieken wel of juist niet zijn toegepast.

Expliciete verificatie door middel van triangulatie met verschillende gegevensbronnen, onderzoekmethoden of onderzoekers bevordert de stabiliteit van de onderzoekuitkomsten en dus hun betrouwbaarheid. Dit is bij de programma-evaluaties in het kader van de medefinanciering in het bijzonder relevant omdat in die onderzoeken vrijwel uitsluitend gebruik is gemaakt van kwalitatieve methoden. Deze laten meer ruimte voor subjectiviteit bij de interpretatie van ruwe gegevens dan kwantitatieve technieken. Immers, bij kwalitatieve onderzoekmethoden is de onderzoeker zelf veel vaker het belangrijkste onderzoekinstrument. Bovendien is het, gegeven de aard van de evaluatieobjecten onontkoombaar dat het grootste deel van de geraadpleegde documentatie en informanten afkomstig is of deel uitmaakt van partijen die een reëel – zij het legitiem – belang hebben bij de uitkomsten van het evaluatieonderzoek. Deze omstandigheden maken aandacht voor triangulatie relevant.

Reikwijdte

Bij programma-evaluaties kan doorgaans maar een klein deel van de individuele activiteiten of projecten die samen het onderzoekobject vormen, daadwerkelijk worden onderzocht. Omdat deze studies worden ondernomen om verantwoording af te leggen en/of lessen te trekken op beleidsniveau, is het van belang vast te stellen in hoeverre de bevindingen van de onderzochte gevallen gelden voor het *hele* evaluatieobject, inclusief de niet bestudeerde gevallen. Deze zogenoemde externe validiteit hangt af van de *representativiteit* van de bestudeerde steekproef voor de totale (operationele) populatie waarover het onderzoek geldige uitspraken wil doen. Veel onderzoeken van MFS-organisaties hebben betrekking op programma's die een groot aantal afzonderlijke activiteiten omvatten en een sterke geografische spreiding vertonen, bijvoorbeeld over een heel continent of zelfs de hele Derde Wereld. In die gevallen is het ter plaatse onderzoeken van een aselecte steekproef van een voldoende omvang om voldoende representatief te kunnen zijn voor een hele populatie, om logistieke en financiële redenen onhaalbaar. Men zal dan de ambitie van

statistische representativiteit moeten opgeven en noodgedwongen kiezen voor een beperkt aantal casestudies. Echter ook in dat geval kan men aandacht besteden aan de (beperkte) generaliseerbaarheid van bevindingen en de reikwijdte van conclusies. Het is dan ook zaak de selectie van de onderzochte gevallen juist in dit licht zorgvuldig en expliciet te beargumenteren zodat ook buitenstaanders zich er een mening over kunnen vormen.

Weinig factoren zijn bevorderlijker voor de geloofwaardigheid van evaluatie-uitkomsten dan de openhartige vermelding van de *beperkingen* waaronder het onderzoek is uitgevoerd en, vooral, van de consequenties die hieruit moeten worden getrokken voor de bruikbaarheid en het gebruik van de onderzoekresultaten.

Onafhankelijkheid

Om praktische en financiële redenen zijn de mogelijkheden voor primaire gegevensverzameling doorgaans beperkt. Vandaar dat de meeste PE's in belangrijke mate steunen op secundair materiaal. Het is welhaast onvermijdelijk dat dergelijk materiaal overwegend afkomstig is uit *bronnen* die nauw aan de te onderzoeken programma's zijn gelieerd. Van evaluatoren mag worden verwacht dat zij hiermee rekening houden, vooral door actief te zoeken naar mogelijkheden om ook materiaal uit volledig onafhankelijke bronnen in hun onderzoek te betrekken.

Voor de uitvoerders van programma-evaluaties geldt dat die nooit helemaal onafhankelijk van alle belanghebbenden kunnen zijn, omdat zij uiteindelijk door de betrokken MFS-organisaties als opdrachtgevers moeten worden gecontracteerd, geïnstrueerd, gedechargeerd en betaald. Met inachtneming van die onvermijdelijke beperking kan er wel op worden toegezien dat de *onderzoekers* in ieder geval geheel onafhankelijk staan ten opzichte van het evaluatieobject. Dat wil zeggen dat zij geen eerdere bemoeienis met het te evalueren programma hebben gehad. De noodzakelijke distantie ten opzichte van het evaluatieobject kan er verder mee gediend zijn als bij de rekrutering – in ieder geval van de teamleider – aan evaluatie-expertise voorrang wordt gegeven boven materiedeskundigheid.

Onderzoekverloop

Voor een goede beoordeling van de onderzoekuitkomsten dient de rapportage een verantwoording te bevatten van het *verloop* van de evaluatie. Daarbij is het vooral van belang de eventuele afwijkingen van de oorspronkelijke opzet, zoals vastgelegd in de ToR, duidelijk weer te geven en te beargumenteren

De uitvoering van een programma-evaluatie dient, nadat deze is uitbesteed aan externe evaluatoren, nauwlettend door de opdrachtgever te worden gevolgd. Alleen dan kan tijdig worden bijgestuurd of anderszins worden ingegrepen in de onderzoekopzet of –uitvoering als onvoorziene ontwikkelingen daar aanleiding toe geven. Veelal is het nuttig voor de supervisie van de PE een *ad hoc* begeleidings- of stuurgroep in het leven te roepen, bestaande uit interne en/of externe belanghebbenden en deskundigen. Het is niet verstandig het voorzitterschap van een dergelijk gezelschap in handen te leggen van een functionaris die direct bij het geëvalueerde programma is betrokken (geweest).

4.2.3 Bruikbaarheid

Voor zover de bruikbaarheid niet afhangt van de validiteits- en betrouwbaarheidscomponenten beperkt de verdere concretisering van dit criterium zich tot een tweetal praktische indicatoren, elk bestaande uit eveneens twee beoordelingscomponenten.

Presentatie

Voor een effectief gebruik van PE-resultaten is het uiteraard van belang dat de *gebruiksdoelstelling* van het onderzoek ondubbelzinnig is geformuleerd en alle betrokkenen helder voor ogen staat. Dit voorkomt onproductieve en soms pijnlijke misverstanden over de reden(en) waarom een evaluatie is ondernomen en waarvoor de uitkomsten zullen worden gebruikt. Met name is het van belang de (externe) gebruiks- of onderzoekdoelstelling duidelijk te onderscheiden van de probleemstelling van de evaluatie zelf. Gezien de verwarring die hierover in de praktijk frequent kan worden aangetroffen, is dit aspect van meer dan theoretisch belang.

Een randvoorwaarde voor effectief gebruik van onderzoekresultaten is uiteraard dat ze voor de beoogde gebruiker *toegankelijk* zijn. Dat betekent niet alleen dat de rapportage duidelijk en volledig hoort te zijn, maar ook dat ten behoeve van de informatievoorziening aan en besluitvorming door het management de essentie van het onderzoek, en vooral de hoofdbevindingen, kernachtig zijn samengevat.

Aansluiting

Voor opdrachtgevers en andere 'afnemers' van PE-onderzoekresultaten hangt de bruikbaarheid, meer dan van iets anders, af van de mate waarin de uitkomsten beantwoorden aan hun specifieke (informatie)behoeften. Daarbij gaat het er in de eerste plaats om of de conclusies van een evaluatie de *onderzoekvragen* waarin de probleemstelling werd geoperationaliseerd, beantwoorden en hoe volledig zij dat doen. Dit laatste hangt overigens niet alleen af van de uitvoerder van het onderzoek. Ook de opdrachtgevers kunnen er een positieve invloed op uitoefenen door te voorkomen dat het onderzoekontwerp overbelast raakt als gevolg van het al eerder gesignaleerde probleem van de overdaad aan onderzoekvragen.

Voor de opdrachtgevers en andere gebruikers van evaluatieonderzoek vormen de *aanbevelingen* vaak het meest bruikbare onderdeel van programma-evaluaties. Uiteraard dienen de aanbevelingen in de eerste plaats logisch te volgen uit de bevindingen en conclusies van het evaluatieonderzoek. Dit blijkt in de evaluatiepraktijk minder vanzelfsprekend dan men wellicht zou verwachten.Verder moet de implementatie van de gepresenteerde aanbevelingen binnen het bereik van de programma-verantwoordelijken liggen. Anders is er immers voor hen geen sprake van bruikbare stuurinformatie.

4.3 Waarderingsmethode

De beoordelingslijst opent met de registratie van een aantal identificerende basiskenmerken van de betreffende programma-evaluatie. Deze kenmerken hebben achtereenvolgens betrekking op het onderzoeksobject van de beoordeelde PE, het uitgevoerde onderzoek zelf, de bij het onderzoek betrokken actoren en de rapportage over het evaluatieonderzoek. Ze spelen geen rol bij de beoordeling van de methodologische kwaliteiten.

Vervolgens vindt de kwaliteitstoetsing plaats aan de hand van de drie criteria validiteit, betrouwbaarheid en bruikbaarheid door beoordeling op elk van de 20 componenten. Voor elke component wordt een waardering is toegekend op een vierpuntsschaal met de scoremogelijkheden: *slecht – matig – redelijk – goed*. De keuze voor een korte schaal met een even aantal scoremogelijkheden was bewust, waarbij de volgende overwegingen golden:

- Een korte schaal, met slechts vier scoremogelijkheden, vermijdt de pretentie dat de beoordelaar met een nauwkeurigheid van meer dan 25% het ene onderzoek – op elk van de beoordelingsaspecten – van het andere zou kunnen onderscheiden. Een langere schaal, zoals de in het onderwijs veel gebruikte 10-puntsschaal, voert bij de beoordeling van evaluatieonderzoek makkelijk tot schijnnauwkeurigheid.
- Met een even aantal scoremogelijkheden kent de schaal verder een duidelijk omslagpunt – in dit geval tussen de tweede en derde score – dat de beoordelaar dwingt tot kiezen. Dit voorkomt dat, bijvoorbeeld uit gemakzucht, de "veilige" centrumscore wordt opgezocht, zoals bij een oneven schaal wel mogelijk is. In de praktijk wordt de 'neutrale' centrumscore op oneven schalen nogal eens gebruikt voor 'weet niet', 'geen mening' of omdat de respondent het gevoel heeft bij die keuze minder uit te hoeven leggen dan bij een meer geprononceerde score.

De manier waarop de verschillende scores – slecht, matig, redelijk en goed – worden toegekend, is als gevolg van het kwalitatieve karakter van vrijwel alle beoordelingscomponenten en de sterke variatie in beoordelingsobjecten (de PE's) onmogelijk te objectiveren. Individuele beoordelingen moesten dus van geval tot geval worden vastgesteld, waardoor behoefte ontstond aan normering van de toe te kennen scores. Hiertoe werden de vuistregels gehanteerd die in Figuur 4–2 zijn samengevat.

Figuur 4–2 *Vuistregels voor scoretoekenning*

'slecht' is over het algemeen als waardering toegekend als een beoordelingscomponent door de onderzoeker volledig is genegeerd. Omdat alle componenten in de lijst relevant worden geacht voor de kwaliteit van een evaluatieonderzoek, leidt het ontbreken ervan per definitie tot de waardering slecht. Zij werd verder gehanteerd wanneer uit het onderzoek bleek dat de interpretatie van de strekking van de betreffende component ongerijmd was.[33]

'matig' is als kwalificatie gehanteerd in gevallen waarin aan een component niet meer dan rituele aandacht is besteed. Het betreffende aspect wordt dan wel vermeld, maar zonder dat er verder iets mee wordt gedaan dat in de uitwerking of uitkomsten van het onderzoek duidelijk herkenbaar is. In het spraakgebruik wordt hiervoor wel het beeld van 'de klok en de klepel' gehanteerd.

'redelijk' is gebruikt als score gebruikt in situaties waarin aan een onderzoekaspect zodanig inhoudelijk aandacht is besteed, dat het positieve effect daarvan zichtbaar is in de uitkomsten van het beoordeelde evaluatieonderzoek. Dit betekent niet dat in deze gevallen de perfectie is bereikt of dat er niets meer te verbeteren valt, maar wel dat de bedoelde doorwerking aanwijsbaar is en als (ruim) voldoende valt te kwalificeren.

'goed' werd als oordeel gegeven wanneer het betreffende onderzoekaspect zodanig was ingevuld dat daaraan in de gegeven omstandigheden (vrijwel) niets meer te verbeteren viel. Daarbij ging het dus niet om de theoretisch bereikbare volmaaktheid, maar om wat binnen de realiteit van de onderzoekssituatie, in redelijkheid maximaal haalbaar was. Dit betekende dat rekening is gehouden met onvermijdelijke financiële en logistieke beperkingen die bijvoorbeeld voortvloeiden uit de aard van de geëvalueerde interventies of de situatie in het betreffende land, enz.

[33] Uiteraard was de Beoordelingslijst Programma-evaluaties openbaar. In de meeste gevallen werden de gecontracteerde externe evaluatoren dan ook door hun opdrachtgevers bij de MFS-organisaties van tevoren op het bestaan en de functie van de lijst geattendeerd, opdat zij er bij de uitvoering van hun onderzoekopdracht rekening mee zouden kunnen houden.

Hoe korter de waarderingsschaal, hoe groter uiteraard de kans om bij de oordeelsvorming *tussen* in plaats van precies *op* de gegeven scoremogelijkheden uit te komen. Dergelijke situaties hebben zich ook in deze studie voorgedaan. Als het niet mogelijk bleek door de verzameling van extra gegevens de keuze voor de ene of de andere score sterker te onderbouwen, werd door selectie van de hoogste score het oordeel naar boven afgerond. In die gevallen genoten de belanghebbenden dus het voordeel van de twijfel.

Om de waarderingen voor de afzonderlijke beoordelingscomponenten te kunnen combineren tot scores op het niveau van de indicatoren of, nog verder, dat van de drie kwaliteitscriteria, moet met die waarderingen kunnen worden gerekend. Dat betekent in de eerste plaats dat de nu gebruikte ordinale waarderingsschaal moet worden geïnterpreteerd op een hoger meetniveau (interval- of ratioschaal) dan waarop deze bij de toekenning van de scores is gehanteerd. En vervolgens dat de relatieve gewichten van de individuele componenten of indicatoren ten opzichte van elkaar dienen te worden vastgesteld op een zodanige manier dat het resultaat voor een bonte variëteit aan evaluatie-objecten en organisaties toepasbaar en verdedigbaar is.

Met name de tweede voorwaarde bleek een onoverkomelijk obstakel te vormen. Hoewel alle 20 beoordelings–componenten relevant kunnen worden geacht voor de kwaliteit van PE-onderzoek, is het onwaarschijnlijk dat ze allemaal van precies *evenveel* belang zijn voor het kwaliteitsoordeel. Reden dus om tussen de verschillende beoordelingscomponenten te differentiëren. Helaas was de bepaling en onderbouwing van relatieve gewichten al tijdens de MFP-breed subsidieperiode onhaalbaar gebleken, omdat geen solide theoretische basis voor de differentiatie van gewichten kon worden gevonden. Daardoor behield elke poging tot gewichtsverdeling een zodanig arbitrair karakter, dat dit meer discussie losmaakte dan de toekenning van de beoordelingsscores zelf. Vanwege de onmogelijkheid een bruikbare oplossing voor dit lastige vraagstuk te vinden is van verdere pogingen tot ontwikkeling van samengestelde scores afgezien.

Overigens zijn de gevolgen voor het gebruiksgemak van de vervaardigde beoordelingen, zeker ten behoeve van de leerfunctie beperkt. Immers, met het advies om – in algemene zin – de validiteit of betrouwbaarheid van bepaalde onderzoeken tegen het licht te houden en te verbeteren, kan een evaluatiemanager in de praktijk weinig. Adviezen en maatregelen voor methodologische verbeteringen zijn pas echt uitvoerbaar als ze voldoende concreet geformuleerd zijn. En dat is een stuk eenvoudiger op het lagere aggregatieniveau van de beoordelingscomponenten (bijvoorbeeld 'de steekproefverantwoording') te doen dan op het hogere abstractieniveau van bijvoorbeeld het validiteitscriterium.

4.4 Beoordelingsprocedure

Alle PE-beoordelingen doorliepen verscheidene fasen voordat zij resulteerden in vastgestelde kwaliteitsoordelen.

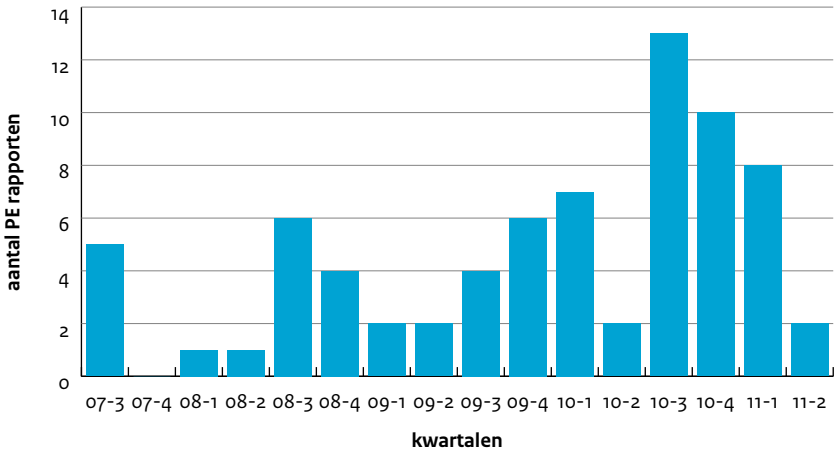
4.4.1 Aanlevering van studies

Het beoordelingstraject begon met de aanlevering van het eindrapport van de betreffende programma-evaluatie bij IOB. De aanlevering van rapporten door de elf organisaties, zoals weergegeven in Figuur 4–3, verschilde weinig van het patroon dat zich manifesteerde gedurende de periode 2003 – 2006, dat wil zeggen het begon met een langzame aanloop en vertoonde tegen het einde een piek. Dit ondanks het feit dat de aanbieding van PE-studies ter beoordeling door de vier grote MFO's, die uit het MFP-breed programma afkomstig waren, gelijkmatiger verliep.

Bij de zeven overige MFS-subsidieontvangers, voor wie deze vorm van toezicht nieuw was, kwam de presentatie van de onderzoeksrapporten echter moeizaam op gang. Dit was vooral toe te schrijven aan de onduidelijke manier waarop van tevoren door het ministerie met de betrokken organisaties was gecommuniceerd over wat er van hen met betrekking tot de uitvoering van programma-evaluaties werd verwacht. Zo werden de elf betrokken organisaties voor het eerst over reikwijdte en inhoud van het toezicht op de door hen te ondernemen programma-evaluaties geïnformeerd tijdens een Monitoring & Evaluatie (M&E) werksessie die het ministerie op 16 oktober 2006 organiseerde. Dit was ruim drie weken *nadat* de besluiten over de subsidietoekenningen bekend waren gemaakt. De bijeenkomst was bestemd voor alle 58 organisaties waaraan op 22 september een subsidie was toegekend en, mede daardoor, voornamelijk gewijd aan de introductie van het ingrijpende en bewerkelijke systeem van *Maatgesneden Monitoring* (MM)[34] dat voor *alle* grote én kleine subsidieontvangers met ingang van 1 januari 2007 zou gaan gelden. Vergeleken daarmee was de invulling van het toezicht op de kwaliteit van programma-evaluaties, dat slechts op een beperkt deel van de vertegenwoordigde organisaties van toepassing zou zijn, een zaak van ondergeschikt belang.

[34] Zie voor een 'beperkte beleidsdoorlichting' evaluatie van het MM-systeem IOB-rapport no. 321: *Maatgesneden Monitoring 'Het verhaal achter de cijfers'*, april 2009. Deze studie concludeerde onder meer dat de zware administratieve last die het MM-systeem voor de beperkte beheercapaciteit van het ministerie met zich meebracht niet resulteerde in een evenredige verbetering van het inzicht in door de MFS-organisaties behaalde resultaten. Daartoe werd aanbevolen in de toekomst meer aandacht te besteden aan evaluaties.

Figuur 4–3 *Aanlevering PE rapporten per kwartaal 2007-2011*



Bij de zeven 'grote subsidieontvangers' die niet eerder met (de beoordeling door IOB van) programma-evaluaties te maken hadden gehad, wekte de aankondiging van het toezicht wel de nodige verbazing. Vooral omdat zij ter plaatse werden geconfronteerd met een 'verwachting' van het ministerie waarover zij in de reeds door hen ontvangen formele subsidiebeschikkingen (gedateerd op 29 september 2006) niets hadden gelezen. Omdat het toezicht op de programma-evaluaties niet in deze beschikkingen was opgenomen, kon ook niet worden gesproken van een *verplichting* in juridische zin en moest worden volstaan met het uitspreken van verwachtingen. Tegen deze achtergrond was het niet onbegrijpelijk dat de aanvankelijke verwondering bij de betrokken 'kleinere' organisaties zich in de loop van de subsidieperiode ontwikkelde tot een zekere irritatie over de gang van zaken (vandaar de in § 3.1 vermelde brief aan DSI). Uiteindelijk manifesteerde deze ergernis zich bij enkele organisaties zich in een merkbare weerstand tegen actieve medewerking aan het door IOB uit te oefenen toezicht.

Begin 2010 zag het ernaar uit dat de aanlevering van PE-studies – overigens niet alleen van de 'kleinere' onder de elf betrokken organisaties – achter zou blijven bij wat nodig was om een redelijke dekkingsgraad, zoals gedefinieerd in § 3.3, te bereiken. Hierop besloot de Directie Sociale Ontwikkeling (DSO), als coördinerende directie binnen het ministerie de opvolger van DSI, aan de bel te trekken. In een serie rappelbrieven aan alle betrokken organisaties werden deze herinnerd aan het belang van de verantwoording van subsidie-uitgaven door middel van programma-evaluaties en aan de in oktober 2006 uitgesproken verwachting dergelijke studies voor een kwaliteitsbeoordeling aan de Inspectie voor te leggen.

Hoewel niet met zekerheid valt vast te stellen of deze actie van doorslaggevende betekenis is geweest, kan wel worden geconstateerd dat zich vanaf medio 2010 een versnelling heeft voorgedaan in de aanlevering van evaluatiestudies. Als gevolg daarvan werd, zoals Figuur 4–3 laat zien, van alle beoordeelde 72 PE's bijna de helft (33) in de laatste vier kwartalen ontvangen.

Om zoveel mogelijk actueel onderzoek in de meta-evaluatie te kunnen betrekken, kon de gegevensverzameling niet precies op 31 december 2010 worden afgesloten. Er werd een korte uitlooptijd in acht genomen om PE's die wel gedurende de subsidieperiode waren uitgevoerd, maar waarvan de rapportage nog niet was voltooid, mee te kunnen nemen. Het laatste van de 72 aangeleverde rapporten werd zodoende ontvangen in april 2011 waarna de beoordeling van de betreffende PE in mei is voltooid.

4.4.2 Concept-beoordeling

Daar het, gegeven de capaciteitsbeperkingen bij de Inspectie, niet haalbaar was om meer dan één inspecteur/evaluator voor het beoordelingsproces in te zetten, moest het risico van eenzijdige oordeelsvorming door IOB op een andere manier worden beperkt. Hiertoe is een uitvoerige beoordelingsprocedure in twee ronden – een schriftelijke en een mondelinge – gevolgd met de bedoeling aan de betreffende MFS-organisaties maximale ruimte te bieden voor commentaar, aanvullingen en correcties. Dit vanuit de gedachte dat zij als verantwoordelijke opdrachtgevers het meeste belang hebben bij de voorkoming en correctie van onzorgvuldige en/of onjuiste beoordelingen van door hen voorgelegde onderzoeken. Bovendien beschikken zij vanuit hun directe en actieve betrokkenheid bij de opzet en uitvoering van hun PE's over de meeste en beste informatie om toe te zien op nauwkeurige en faire beoordelingen.

Na ontvangst van de rapporten bij IOB werden ze eerst gecontroleerd op volledigheid van de informatie die nodig was om een concept-beoordeling te kunnen opstellen. Ontbrekende gegevens werden vervolgens bij de betreffende MFS-organisatie of, in het geval van een gezamenlijke PE, bij de evaluatiemanager van Partos opgevraagd. In vrijwel alle gevallen bleek dit ook nodig. Daarnaast kwam het regelmatig voor dat de beschikbare rapportage op onderdelen te summier was om zonder aanvullende gegevens door buitenstaanders (zoals IOB) volledig te kunnen worden begrepen. Ook in die situaties moest er om aanvullend materiaal worden gevraagd.

Het gevolg was in de meeste gevallen dat er aanzienlijke vertragingen optraden in het beoordelingsproces, waarbij de organisaties in een aantal gevallen opnieuw contact moesten zoeken met de externe onderzoekers die voor de betreffende studie waren gecontracteerd. Hierdoor konden de doorlooptijden van de PE-beoordelingsprocedures behoorlijk oplopen, waarbij perioden van zes maanden en meer geen uitzondering vormden.

Na ontvangst van de aanvullende informatie stelde IOB een voorlopige beoordeling op door de standaard-beoordelingslijst in concept in te vullen. Hierbij werd voor elke component-score via paginaverwijzingen aangegeven op welke passages in het rapport of later ontvan-

gen aanvullingen de betreffende waardering primair (zij het niet uitsluitend) berustte. De voorlopige beoordeling werd vervolgens aan de indiener van het betreffende rapport toegezonden, vergezeld van het aanbod de gegeven scores en de overwegingen waarop zij waren gebaseerd mondeling toe te lichten en eventuele verschillen van inzicht over specifieke waarderingen te bespreken.

4.4.3 Bespreking en vaststelling

In vrijwel alle gevallen gingen de organisaties op dit aanbod in. Eigenlijk bleven dergelijke besprekingen van concept-beoordelingen alleen achterwege als zij door praktische omstandigheden werden verhinderd. Op deze manier vonden bijna 70, doorgaans uitvoerige discussies plaats, waaraan meestal niet alleen evaluatievertegenwoordigers deelnamen, maar ook programmabeheerders die de inhoudelijke verantwoordelijkheid droegen voor de geëvalueerde beleidsonderdelen. Dit bevorderde de diepgang van de gedachtewisseling, omdat deze functionarissen, die meestal de opdrachtgevers voor de beoordeelde PE's waren geweest, van alle achtergronden van onderzoeksopzet en -verloop op de hoogte waren. In enkele gevallen namen, op voorstel van de evaluatievertegenwoordigers van de MFS-organisaties tevens externe onderzoekers die de beoordeelde PE's hadden uitgevoerd aan de bespreking deel.

Behalve voor toelichting op de concept-scores waren de besprekingen, gegeven het legitieme belang van de betrokken MFS-organisaties bij de uitkomst van de IOB-beoordelingen, ook expliciet bedoeld om de vertegenwoordigers van deze organisaties ruim de gelegenheid te bieden om tegengas te geven. Van de mogelijkheden daartoe is – terecht – vrij intensief gebruik gemaakt om nadere mondelinge onderbouwing te verlangen van iedere waardering waarvoor men de overwegingen onjuist of onvoldoende steekhoudend achtte. Dit dwong de Inspectie niet alleen te rechtvaardigen waarom ter discussie gestelde scores precies waren toegekend, maar ook om aan te geven wat opdrachtgevers en onderzoekers in gegeven omstandigheden hadden moeten doen om aanspraak te kunnen maken op een hogere waardering. In alle gevallen waarin de onderbouwing voor een concept-score niet houdbaar bleek tegenover de argumentatie van de kant van de MFS-organisaties, werd de waardering herzien. Dit kwam regelmatig voor.
Na afronding van deze discussie en het aanbrengen van eventueel noodzakelijke wijzigingen werden ten slotte de beoordelingen in definitieve versie vastgesteld.

Alles bijeengenomen was de hele beoordelingsprocedure enigszins omslachtig en, mede daardoor, langdurig. Bij de opzet ervan heeft voor IOB het belang van zorgvuldigheid zwaarder gewogen dan het streven naar snelheid. Voor de MFO's telde, naast het belang dat zij hebben bij zorgvuldigheid, dat de gevolgde werkwijze tijd en gelegenheid bood om uit de beoordelingen lessen te trekken voor toekomstige verbeteringen. In die zin kan het consultatieproces tussen IOB en de MFS-organisaties worden opgevat als een component van de leerfunctie van deze meta-evaluatie. Afgaande op de in het algemeen constructieve opstelling van zowel evaluatie- als programmavertegenwoordigers was de intentie hiertoe bij de betrokkenen onmiskenbaar aanwezig.

5

Onderzoekresultaten

De evaluatie-inspanningen van de elf grote subsidieontvangers onder het MFS-I kunnen zowel vanuit een kwantitatieve als vanuit een kwalitatieve invalshoek worden beschouwd. In het eerste geval gaat het om de omvang van de inspanning, in het tweede om de methodologische deugdelijkheid van de uitgevoerde onderzoeken. In dit laatste hoofdstuk komen beide invalshoeken aan de orde. Eerst de kwantitatieve, dan de kwalitatieve.

5.1 Evaluatiedekking

In § 3.3 werd voor het doel van deze meta-evaluatie het begrip evaluatiedekking gedefinieerd als de verhouding tussen het bedrag aan medefinancieringsmiddelen (in ruime zin) dat door middel van programma-evaluaties wordt verantwoord en het subsidiebedrag dat onder het MFS-I is toegekend. Tijdens de bijeenkomst over Maatgesneden Monitoring (zie § 4.2.1) was aangekondigd dat de dekkingsgraad voor de subsidieperiode 2007 – 2010 ten minste 40% zou dienen te bedragen. Dit betekende voor de betrokken organisaties dat zij zoveel met MF-middelen gefinancierde activiteiten zouden moeten (laten) evalueren, dat op basis daarvan geldige evaluatieve uitspraken mogelijk zouden zijn over een bedrag equivalent aan ten minste 40% van de ontvangen MFS-I subsidie. Deze omslachtige formulering is nodig om aan te geven dat de af te dekken '40%' niet letterlijk en *integraal* behoefde te worden onderzocht. De 'dekkingseis' vergde slechts de evaluatie van een *representatief gedeelte* van een uitgavenvolume dat 40% van de ontvangen subsidie bedroeg. Dan zouden namelijk de bevindingen en conclusies van de onderzochte gevallen geldig geacht kunnen worden voor het grotere aantal niet onderzochte.

Op dit punt deed zich het methodologische probleem voor dat in het kader van de beoordeelde PE's, afgezien van een enkele enquête, nauwelijks kwantitatieve onderzoekmethoden worden toegepast. Hierdoor kan er in veruit de meeste gevallen van representativiteit in statistische zin geen sprake zijn. De dominante onderzoekmethode in bijna alle programma-evaluaties is daarentegen de casestudy, met gebruikmaking van kwalitatieve technieken als dossieronderzoek, (diepte-)interviews en focusgroepen. Om te vermijden dat de uitgevoerde casestudies alleen representatief zouden zijn voor zichzelf was het van groot belang dat hun selectie zorgvuldig plaatsvond en werd beargumenteerd. Veel van die discussies tussen IOB en vertegenwoordigers van de MFS-organisaties hadden betrekking op dit aspect. Niet alleen omdat het een methodologisch lastig onderdeel van een goede PE-opzet vormde, maar ook omdat de representativiteit van doorslaggevende invloed kon zijn op de evaluatiedekking van het betreffende onderzoek. En daarmee op de bijdrage van de betreffende PE aan de te bereiken (totale) dekkingsgraad van 40%.

In Tabel 5–1 zijn alle 72 programma-evaluaties opgenomen met het uitgavenbedrag waarmee ze bijdragen aan de bereikte evaluatiedekking (in euro's). Met 47% van de € 1,65 miljard die aan de elf organisaties samen aan MFS-subsidie werd toegekend, is over het geheel genomen de minimum evaluatiedekking van 40% ruimschoots bereikt. Ook individueel hebben de elf organisaties weinig moeite met het drempelpercentage, waarbij enkele, die in feite uit eigen beweging streven naar een 100% dekking, dit zelfs ver overtreffen. De enige duidelijke uitzondering is IKV/Pax Christi dat zich onder andere van de overige

organisaties onderscheidt doordat het een groot aantal relatief kleinschalige programma's kent, die alleen afzonderlijk van elkaar zinvol kunnen worden geëvalueerd. Dit heeft tot gevolg dat deze organisatie in vergelijking met de overige een grotere evaluatie-inspanning – uitgedrukt in de verhouding tussen evaluatie-uitgaven en programmakosten – moet plegen om eenzelfde evaluatiedekking te bereiken (zie ook § 5.2).

Tabel 5–1 Evaluatiedekking Programma-evaluaties

PE titel	#	PE-uitgaven	totaal geëvalueerd	subsidiebedrag	dekking (%)
CORDAID	**11**		**151.249.348**	**396.736.000**	**38**
Common Programming and Shared Indicators		800.000			
PHC Support Programme in Cyangugu Province		2.500.000			
Biologische Fruitteelt Latijns-Amerika		2.300.000			
Kinderprogramma		4.600.000			
Midden-Oosten en Midden- en Oost-Europa		9.667.000			
Capaciteitsversterking Financieel Management		2.600.000			
Sloppenwijkbewoners		7.241.037			
Performance-Based Financing		3.804.556			
SME Support an Value Chain Development		18.000.000			
Disaster Risk Reduction		12.830.713			
HIV Programma		19.990.000			
Draagvlakonderzoek (IOB)		5.024.560			
Nicaragua Landenstudie (IOB)		3.500.000			
Conflicttransformatie (Partos)		14.700.000			
Microfinanciering (Partos)		37.391.354			
Inheemse Volken (Partos)		6.300.128			
HIVOS	**7**		**113.231.074**	**231.995.000**	**49**
LGBT Emancipation in Southern Africa		2.842.000			
Missing Links: Organic Chains		6.900.000			
Evaluación del Programa Arte y Cultura		3.200.000			
Tsunami Response		2.878.221			
Human Rights and Democratisation in Kenya		4.300.000			

PE titel	#	PE-uitgaven	totaal geëvalueerd	subsidiebedrag	dekking (%)
ICT & Media Programme		13.542.218			
HIV & AIDS Programme		6.900.000			
Draagvlakonderzoek (IOB)		1.224.228			
Nicaragua Landenstudie (IOB)		11.200.000			
Microfinanciering (Partos)		48.111.935			
Inheemse Volken (Partos)		12.132.472			
ICCO	**9**		**193.454.805**	**443.961.737**	**44**
Stille Oceaan Programma		4.320.000			
Kameroen Programma		3.160.340			
ACT Lobbycampagne Indonesië		52.000			
Strong People, Strong State		7.807.527			
Technical and Vocational Education and Training		9.331.009			
Terrafina Microfinance Mid-Term Review		5.360.000			
International Markets Programme		37.613.605			
Local Market Development Programme		41.124.961			
Food Security Programme		28.603.068			
Draagvlakonderzoek (IOB)		4.893.937			
Nicaragua Landenstudie (IOB)		4.400.000			
Conflicttransformatie (Partos)		5.300.000			
Microfinanciering (Partos)		27.473.363			
Inheemse Volken (Partos)		14.014.995			
OXFAM-NOVIB	**12**		**238.752.538**	**438.946.000**	**54**
Core Country Programme Evaluation Nigeria		4.909.000			
Core Country Programme Evaluation Somalia		8.798.000			
Core Country Programme Evaluation India		17.553.000			
Review Linkis Oxfam Novib - Draagvlakversterking		14.200.000			
Core Country Programme Evaluation Angola		5.835.157			
Humanitarian Capacity Building		5.751.142			

PE titel	#	PE-uitgaven	totaal geëvalueerd	subsidiebedrag	dekking (%)
CCPE Palestinians in the Middle East		7.827.000			
Moving a Mountain - Stop EPA Campaign		884.000			
Core Country Programme Evaluation Ethiopia		16.426.893			
Global Programme		18.152.000			
Core Country Programme Evaluation Burundi		5.023.783			
Evaluación del Programa de América Latina		56.866.000			
Draagvlakonderzoek (IOB)		1.405.900			
Nicaragua Landenstudie (IOB)		6.700.000			
Conflicttransformatie (Partos)		9.100.000			
Microfinanciering (Partos)		59.320.663			
ETC	**2**		**5.152.724**	**12.987.552**	**40**
PROLINNOVA		2.017.000			
ETC Energy Access Evaluation		3.135.724			
ICS	**8**		**11.461.476**	**21.433.634**	**53**
Rijstbanken Programma		322.689			
Action for Children Programme		2.285.000			
Aflatoun Secretariat		2.158.000			
Knowledge Centre on Civic Driven Child Development		680.745			
African Child Policy Forum (ACPF)		1.947.685			
CDCD Programme Cambodia		613.004			
Change for Children (CfC)		2.038.749			
Child Rights in School Improvement Programme		1.415.604			
IKV PAX CHRISTI	**9**		**3.103.876**	**15.090.000**	**21**
Local Peace Initiatives in Ituri, DRC		288.482			
The European Coalition on Oil in Sudan		451.661			
Giving the Syrian Voices of Change More Than Some Extra Oxygen!		334.993			
The Cluster Munition Coalition		60.045			

PE titel	#	PE-uitgaven	totaal geëvalueerd	subsidiebedrag	dekking (%)
United Civilians for Peace (UCP)		125.805			
Kosovo Programme		771.319			
Nagorno Karabach Programma		233.788			
Sudan Integrated Peace building Programme (SIPP)		428.783			
Norte del Cauca Programa		409.000			
IUCN	**1**		**15.770.361**	**25.710.176**	**61**
Ecosystems and Human Well-Being Programme		15.770.361			
SOLIDARIDAD	**3**		**7.064.019**	**16.619.525**	**43**
The impact of Fair Trade		1.216.875			
Food, Feed and Fuel for Thought		2.262.964			
Evaluation Fashion Programme Solidaridad 2003 - 2010		3.584.180			
TERRE DES HOMMES	**4**		**25.515.676**	**32.000.000**	**80**
Pathuakali Development Project (PDP) & Coastal Island Health and Education Project (CIHEP)		2.193.216			
Mathagondapalli Education Centre (MEC)		422.738			
Southeast Asia Campaign against Child Trafficking		3.175.130			
Onderwijs- en Gezondheidprogramma		19.724.592			
WOORD & DAAD	**3**		**18.362.837**	**19.218.872**	**96**
Education Programme		1.495.378			
Job and Income Programme		7.172.578			
Basic Needs Programme		9.694.881			
PARTOS	**3**				
Conflict Transformation: a Science and an Art					
Microfinanciering					
Inheemse Volken					
Totaal	**71**		**783.118.734**	**1.654.698.496**	**47**

De MF-uitgaven die bestreken werden door de in Partos-verband gezamenlijk uitgevoerde evaluaties zijn uitgesplitst en toegerekend aan de deelnemende organisaties. Dit betreft dus de PE's *Conflicttransformatie*, *Microfinanciering* en *Inheemse Volken* en de (voormalige) MFO's Cordaid, Hivos, ICCO en Oxfam Novib. Aan de evaluatiedekking van deze zelfde vier organisaties zijn verder de bedragen toegevoegd die tijdens de subsidieperiode het voorwerp hebben gevormd van onafhankelijke IOB-studies, namelijk het *Draagvlakonderzoek* uit 2009 en de *Evaluatie van de activiteiten van de Medefinancieringsorganisaties in Nicaragua* uit 2010.

5.2 Algemene PE-kenmerken

Hoewel de kenmerken van het evaluatieobject, het evaluatieonderzoek, de actoren en de rapportage waarmee de beoordelingslijst opent, zelf geen rol spelen bij de kwalitatieve beoordeling zijn ze wel relevant voor het algemene beeld van de bestudeerde PE's. Een aantal daarvan is daarom opgenomen in bijlage 2. De meeste van de geregistreerde kenmerken spreken zichzelf. Een enkele behoeft nadere toelichting.

Onderzoekdoelstellingen

Voor de MFS-organisaties zijn programma-evaluaties meer dan andere vormen van M&E (mede) gericht op het afleggen van verantwoording. De meeste PE's vermelden dan ook in hun onderzoekontwerp onderzoekdoelstellingen zowel de *verantwoordingsfunctie* als de *leerfunctie*. Dat neemt niet weg dat bij ruim een derde (26) van de programma-evaluaties de externe doelstelling van het onderzoek uitsluitend wordt geformuleerd in termen van 'leren', 'lessen trekken' en 'beleidsontwikkeling'. In slechts één geval staat (impliciet) de zogenaamde *managementfunctie* voorop in de zin dat het onderzoek voornamelijk is geëntameerd ter voorbereiding van de besluitvorming over de toekomst van een activiteit (*United Civilians for Peace* – IKV/Pax Christi).

Onderzoekuitgaven

Het financiële beslag van de PE-objecten, in termen van beslag op MF-middelen, was niet altijd meteen duidelijk. Een groot deel van de betrokken organisaties moest eraan wennen dat het hierbij ging om meer dan om een administratief detail. Tegen het einde van de subsidieperiode was het belang van dit gegeven voor de bepaling van de 'evaluatiedekkingsgraad' echter overal doorgedrongen. Geholpen door de in § 4.2.1 genoemde herinneringsactie van DSO had dit aangescherpte besef waarschijnlijk een stimulerend effect op de aanlevering van PE-studies ter beoordeling.

In totaal bestreken de 72 beoordeelde programma-evaluaties €756 miljoen aan MF-uitgaven, met een gemiddelde omvang van € 10,5 miljoen. Achter dit gemiddelde verbergt zich een brede spreiding, zoals Tabel 5–1 laat zien: van enkele tienduizenden tot enkele tientallen miljoenen euro's, met als uitschieter de gezamenlijk PE *Microfinanciering* waarmee een beslag op MF-middelen van € 172 miljoen gemoeid was. De overige twee studies in Partos-verband, *Conflicttransformatie* en *Inheemse Volken*, bevonden zich met respectievelijk € 29 miljoen en € 32 miljoen qua financieel belang ook in de hogere regionen.

Met onderzoekuitgaven wordt hier alleen gedoeld op de zogenaamde *out of pocket expenses* ten behoeve van evaluatieonderzoek. Deze uitgaven zijn additioneel aan de eventuele 'vaste' kosten voor evaluatiecapaciteit binnen de MFS-organisatie, die deel uitmaken van de overhead en doorgaans voornamelijk bestaan uit loonkosten van eigen evaluatiemedewerkers. Evenals de geëvalueerde bedragen variëren de uitgaven voor evaluatieonderzoek sterk, niet alleen in absolute zin, maar ook in verhouding tot de financiële omvang van de geëvalueerde activiteiten. Uitschieters naar boven in absolute bedragen vormen, zoals te verwachten, de gezamenlijke Partos-evaluaties die uitgaven van € 338.227 tot 454.852 vergden. De uitgaven ten behoeve van individuele PE's liepen uiteen van € 0, in die gevallen waarin een onderzoek helemaal als bureaustudie en door eigen personeel werd uitgevoerd, tot € 229.159 voor een wereldomspannende studie van *lobby and advocacy*-campagnes.

Desgevraagd zal iedere subsidieontvanger verklaren 'groot belang' te hechten aan serieus evaluatieonderzoek. Om aan dit 'beleden' belang enig reliëf te geven, is geïnventariseerd hoeveel de verschillende organisaties uitgaven aan de in hun opdracht ondernomen programma-evaluaties. Omdat de relatieve hoogte van de onderzoekuitgaven meer zegt over de evaluatie-inspanning dan de absolute omvang van de bedragen, zijn de evaluatie-uitgaven vervolgens uitgedrukt in percentages van de financiële belangen van de betreffende evaluatie-objecten. Deze percentages variëren, evenals de evaluatie-uitgaven zelf, namelijk van 0 tot circa 20%. Het gemiddelde voor alle 72 PE's kwam uit op 2,4%. De percentages bewegen zich, afgezien van enkele uitschieters van meer dan 3% (waarvan de meeste organisaties er één of twee kennen), tussen 0,5 en 2%. Een opvallende uitzondering hierop wordt gevormd door de organisatie IKV/Pax Christie. Hoewel haar evaluatie*dekking* zoals weergegeven in Tabel 5–1 met 28% achterblijft bij die van de overige organisaties, ligt haar evaluatie-*inspanning* met een gemiddelde evaluatie-/programmakosten verhouding van 5,7% meer dan twee maal zo hoog als het algemene gemiddelde van 2,4%.

Voor zover bekend bestaan er geen algemeen geaccepteerde normen voor de verhouding tussen evaluatie- en programmakosten. Een uniform percentage ligt ook niet voor de hand als het financiële belang van de evaluatie-objecten, zoals in dit onderzoek, uiteenloopt van 52.000 tot 172 miljoen euro. Voor de hantering van normen voor de verhouding tussen evaluatie-uitgaven en programmakosten zou een glijdende schaal waarschijnlijk realistischer zijn. Daarop zou dan de evaluatiebegroting als percentage van de programmakosten kunnen dalen naarmate de (financiële) programma-omvang zou toenemen, en omgekeerd.

Doorlooptijden

Dat de uitvoering van programma-evaluaties veel tijd kost, is behalve aan de meestal aanzienlijke omvang van de onderzoekobjecten, vooral toe te schrijven aan twee factoren. Ten eerste zijn de (externe) onderzoekers als buitenstaanders in beginsel niet vertrouwd met hun onderzoekobject – de nagestreefde onafhankelijkheid vereist dit zelfs – waardoor zij tijd moeten investeren om het voorwerp van evaluatie te leren kennen voordat ze met hun eigenlijke onderzoek kunnen beginnen. Ten tweede worden diezelfde onderzoekers daarbij gehinderd door de kwaliteit van de dossiervorming, die nogal eens te wensen overlaat. Dit manifesteerde zich vaak in onvoldoende toegankelijkheid van noodzakelijke (resultaat)

gegevens en, meer dan eens, zelfs in de volledige afwezigheid ervan. Samen met tijdgebrek vormt dit obstakel het meest gesignaleerde probleem in de PE-rapportages.

Bij de individuele PE's bedroeg de gemiddelde doorlooptijd tien maanden met een minimum van twee maanden en een maximum van tweeënhalf jaar. De maximale lengte van de doorlooptijd bedroeg bij de gezamenlijke PE's in Partos-verband ook tweeënhalf jaar, maar dan bij een minimum van zestien en een gemiddelde van tweeëntwintig maanden.

5.3 Evaluatiekwaliteit

Zoals werd beschreven in hoofdstuk 4 zijn de methodologische kwaliteiten van de 72 PE's beoordeeld aan de hand van drie criteria (validiteit, betrouwbaarheid en bruikbaarheid), onderverdeeld in 10 indicatoren en vervolgens 20 beoordelingscomponenten. Dit resulteerde in 1.440 afzonderlijke scores die alle zijn weergeven in de individuele scoretabellen voor de elf MFS-organisaties. Deze zijn opgenomen in bijlage 5. Omdat voor het doel van deze meta-evaluatie algemene patronen relevanter zijn dan individuele scores op onderdelen, wordt bij de bespreking van de evaluatiekwaliteit een geaggregeerde benadering gevolgd.

Het aantal gezamenlijke PE's in Partos-verband (3) was te gering om deze studies als categorie op een zinvolle manier te onderscheiden van de individueel uitgevoerde programma-evaluaties.

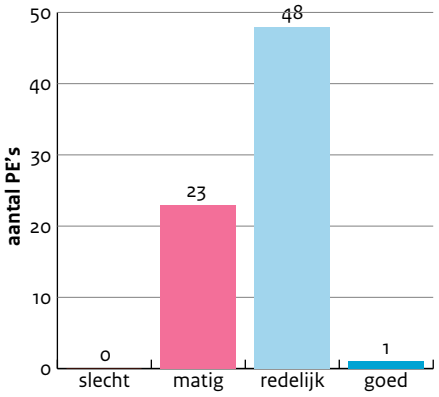
5.3.1 Validiteit

Uit de beoordelingsresultaten blijkt dat de meeste kwaliteitsproblemen zich voordoen bij de validiteit van de onderzochte PE's. Zoals werd toegelicht in § 4.2.1 is deze getoetst aan de hand van vier indicatoren die elk bestonden uit twee componenten. De acht grafieken in deze paragraaf geven de frequentieverdelingen van de waarderingen die voor alle 72 programma-evaluaties aan de afzonderlijke componenten werden toegekend.

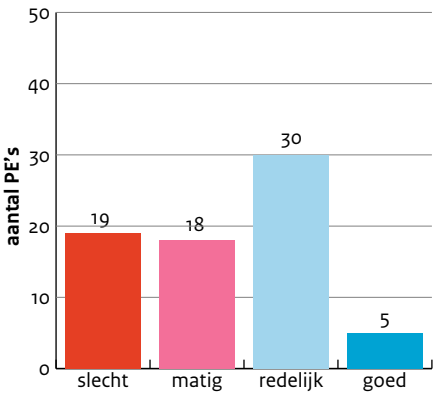
Probleemstelling

Zoals Figuur 5–1 aangeeft, lukt het operationaliseren van de probleemstelling in de meeste gevallen redelijk, maar komen er toch ook nog al eens onduidelijkheden voor. Meestal hebben die betrekking op een te vage uitwerking en formulering van onderzoekvragen, en op overbelasting van de evaluatieopzet met een te groot aantal onderzoekvragen. Als de opdrachtgevers zichzelf in hun ToR in dit opzicht te weinig beperkingen opleggen, kunnen lange reeksen onderzoekvragen ontstaan, waarvan alleen al het aantal het onwaarschijnlijk maakt dat ze (allemaal) worden beantwoord.

Figuur 5–1 *Scores Onderzoekvragen*



Figuur 5–2 *Scores Evaluatiecriteria*

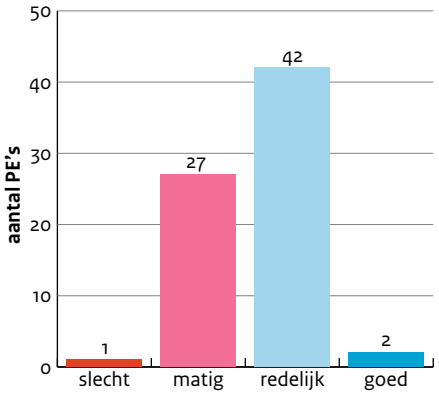


Figuur 5–2 laat zien dat een expliciete en ondubbelzinnige definiëring van de evaluatiecriteria lang niet altijd vanzelfsprekend was. In meer dan de helft van de gevallen lieten de onderzoeken op dit punt veel tot alles te wensen over. Dit probleem is uiteraard betrekkelijk eenvoudig te vermijden. Dat het toch zo vaak voorkomt, lijkt dan ook meer te wijten aan onderschatting van de ernst ervan dan aan onkunde.

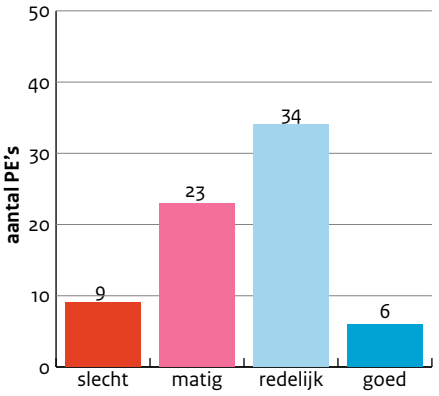
Evaluatieobject

De omschrijving van het evaluatieobject is nagenoeg altijd wel in de een of andere vorm aanwezig, maar vertoont in ruim een derde van de gevallen (Figuur 5–3) de nodige leemten. Het kostte in die gevallen vaak moeite te achterhalen wat precies tot de operationele populatie van de PE behoorde en dus hoe groot het financiële belang ervan was.

Figuur 5–3 *Scores Definitie Evaluatieobject*



Figuur 5–4 *Scores Context Evaluatieobject*

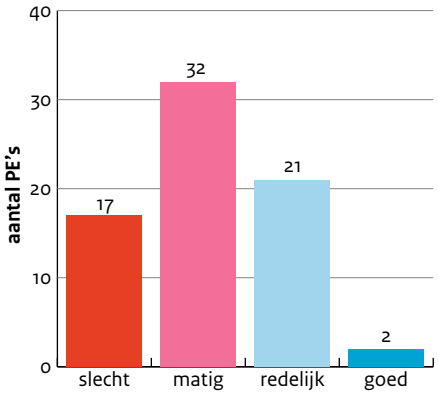


In Figuur 5–4 wordt zichtbaar dat in bijna de helft van de onderzochte gevallen te weinig of helemaal geen werk wordt gemaakt van de beschrijving van de context van het geëvalueerde programma. Vooral de analyse van het institutionele krachtenveld let vaak te wensen over, als ze al niet helemaal ontbrak.

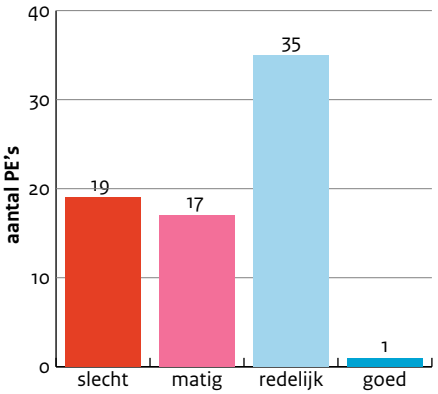
Beleidstheorie

Zoals de scores in Figuur 5–5 en 5–6 duidelijk maken, vormt binnen het criterium van de validiteit de (re)constructie van een coherente en expliciete beleidstheorie het zwakste punt. Dit manifesteert zich niet alleen in de vaagheid of afwezigheid van interventielogica's, maar ook in een onduidelijk onderscheid tussen verschillende resultaatniveaus en in een gebrekkige of ontbrekende operationalisering van abstracte doelen met bruikbare indicatoren.

Figuur 5–5 *Scores Interventielogica*



Figuur 5–6 *Scores Indicatoren*



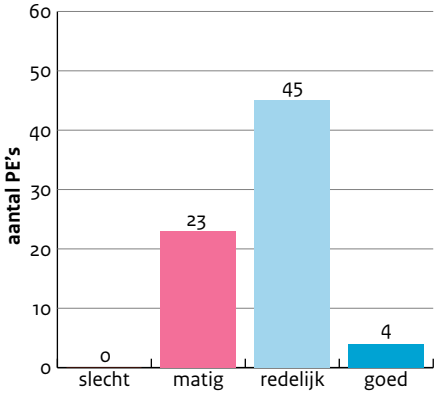
Afgaande op de individuele scores per organisatie en op de PE-beoordelingen uit de MFP-breed periode moet worden vastgesteld dat de tekortkomingen op deze onderdelen niet alleen wijd verbreid zijn maar ook tamelijk hardnekkig. Al kan daarbij wel worden aangetekend dat – mede onder invloed van de aangescherpte eisen van het ministerie op het gebied van monitoring – het *besef* van het belang van goede indicatoren om abstracte resultaten zichtbaar te maken, bij een aantal organisaties wel is toegenomen.

Analyse

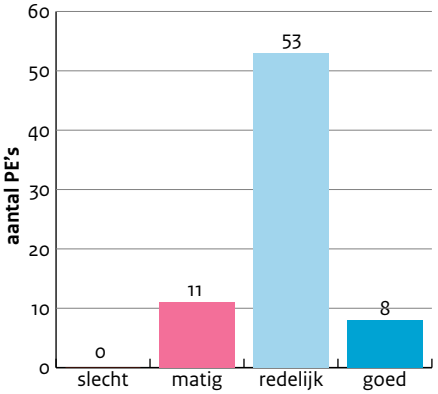
Het beeld van de kwaliteit van de gegevensanalyse is redelijk positief, zoals de scoreverdeling in Figuur 5–7 aangeeft: twee derde van de PE's worden op dit onderdeel als ruim voldoende tot goed beoordeeld. Hierbij past nog de kanttekening dat sommige onderwerpen beperkingen opleggen aan de transparantie waarmee over hun evaluatie kan worden gerapporteerd. Dit doet zich in het bijzonder voor bij activiteiten gericht op het tegengaan van de schending van mensenrechten, waarbij herkenbaarheid van respondenten en informanten in evaluatierapporten voor betrokkenen onaanvaardbare persoonlijke risico's

met zich mee zou brengen. Met name IKV/Pax Christi heeft in een aantal gevallen met deze complicatie te maken.

Figuur 5–7 *Scores Gegevensanalyse*



Figuur 5–8 *Scores Onderbouwing Conclusies*



Figuur 5–8 maakt duidelijk dat de onderbouwing van de gevolgtrekkingen door verzamelde en verwerkte gegevens in verreweg de meeste onderzoeken weinig te wensen overlaat, wat wil zeggen dat ze geen onverklaarbare gedachtesprongen bevatten of conclusies presenteren die niet door het verrichte onderzoek kunnen worden gedragen. Ruim 80% van de PE's scoort op dit onderdeel redelijk tot goed.

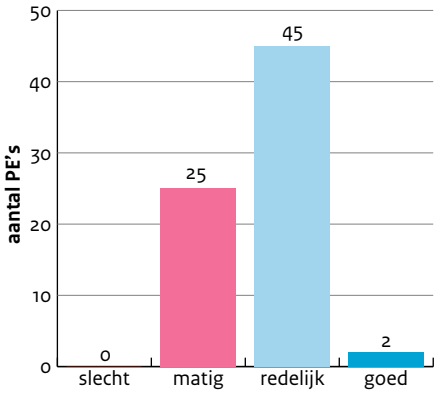
5.3.2 Betrouwbaarheid

Ook de mate waarin de 72 onderzochte PE's voldeden aan dit kwaliteitscriterium is getoetst aan de hand van acht beoordelingscomponenten gegroepeerd in vier indicatoren. De scores op deze componenten zijn weergegeven in de onderstaande acht grafieken.

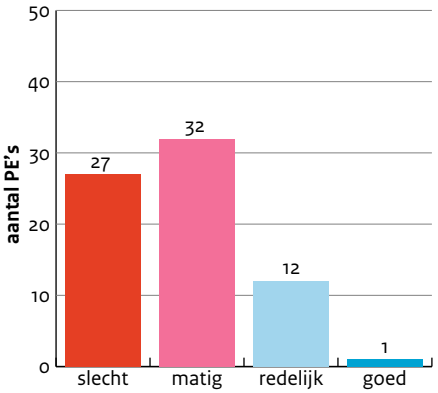
Onderzoekmethoden

Identificatie en beschrijving van de toegepaste onderzoekmethoden zijn over het algemeen redelijk verzorgd, hoewel ook een derde van de PE's, zoals blijkt uit Figuur 5–9, op dit onderdeel als matig is beoordeeld. Dit is dan meestal een gevolg van het feit dat onvoldoende wordt uitgelegd *waarom* bepaalde methoden en technieken wel of juist niet zijn toegepast. Waarbij nog valt aan te tekenen dat bij de bestudeerde PE's vrijwel uitsluitend kwalitatieve onderzoekmethoden zijn gehanteerd. Verder betreffen de waarderingen alleen de specificatie en beargumentering van de gehanteerde onderzoekmethoden en dus niet de vraag of in de gegeven omstandigheden de (meest) geschikte methoden zijn gekozen.

Figuur 5–9 *Scores Onderzoekmethoden*



Figuur 5–10 *Scores Triangulatie*

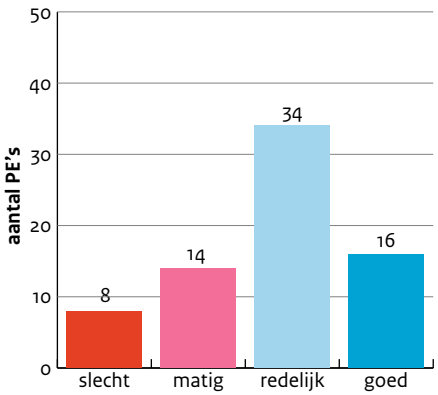


Verificatie door middel van triangulatie met verschillende gegevensbronnen, onderzoekmethoden of onderzoekers is geen gemeengoed (Figuur 5–10). In de PE-rapportages is over het algemeen weinig plaats voor scepsis en controle. Triangulatie zou volgens veel opdrachtgevers van PE's in de praktijk vaker worden toegepast dan uit de dossiers en rapporten blijkt. Als dat klopt, moet het relatief eenvoudig zijn de scorers op dit onderdeel te verbeteren.

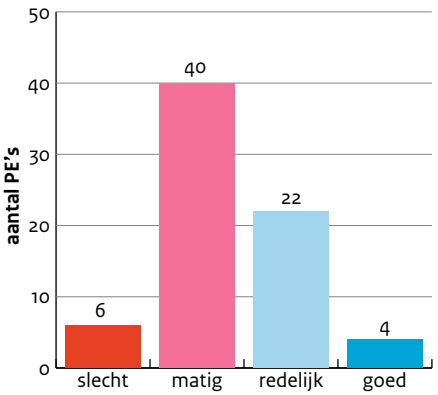
Reikwijdte

Het beeld dat wordt weergegeven in Figuur 5–11 is in zoverre geflatteerd dat bij de meeste als "goed" beoordeelde PE's, deze score in feite berust op de omstandigheid dat hun populaties zo klein waren, dat die in hun geheel konden worden onderzocht. Dan is het dus niet nodig een steekproef te trekken of cases te selecteren en was de representativiteitsvraag niet relevant.

Figuur 5–11 *Scores Representativiteit*



Figuur 5–12 *Scores Beperkingen*



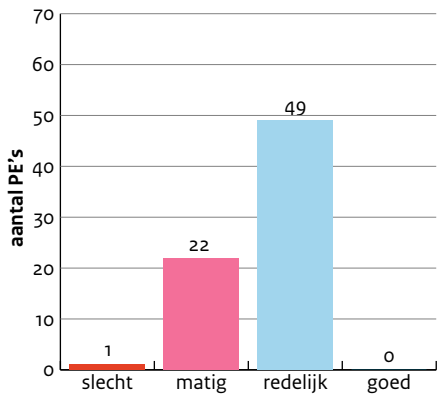
In de meeste PE's beperkte de onderbouwing van de casestudyselectie, als ze niet geheel ontbrak, zich tot factoren als 'gegevensdichtheid', 'diversiteit' en dergelijke. Statistische representativiteit is slechts bij hoge uitzondering en nergens als doel (of probleem) bij het onderzoekontwerp aan de orde gesteld. Het gevolg is dat meestal niet (helemaal) duidelijk wordt hoe generaliseerbaar de onderzoekuitkomsten zijn, overigens zonder dat de opdrachtgevers dit als een obstakel voor de bruikbaarheid van die uitkomsten blijken te ervaren

Waar de opsomming van tegenvallers en obstakels, bij veel PE's nog wel wil lukken, ligt het trekken van conclusies hieruit voor de soliditeit en reikwijdte van de onderzoekbevindingen en dus voor de beantwoording van de onderzoekvragen, kennelijk een stuk moeilijker. Zoals Figuur 5–12 laat zien, bestaat op dit onderdeel nog veel ruimte voor verbetering.

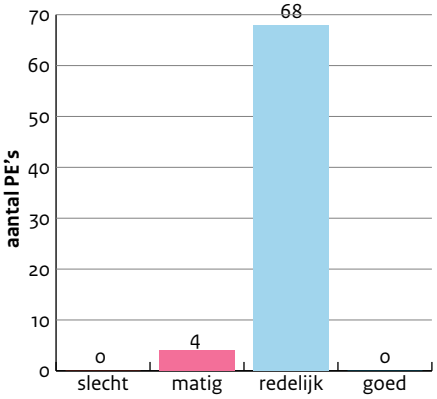
Onafhankelijkheid

Hoewel de hoeveelheid referentiemateriaal vaak indrukwekkend is en belanghebbende partijen als bronnen van informatie meestal onmisbaar, doen sommige evaluatoren wel erg weinig moeite om bronmateriaal buiten te kring van directe belanghebbenden te raadplegen. De "slechte" en 'matige" scores in Figuur 5–13 reflecteren dit.

Figuur 5–13 *Scores Bronmateriaal*



Figuur 5–14 *Scores Onderzoekers*



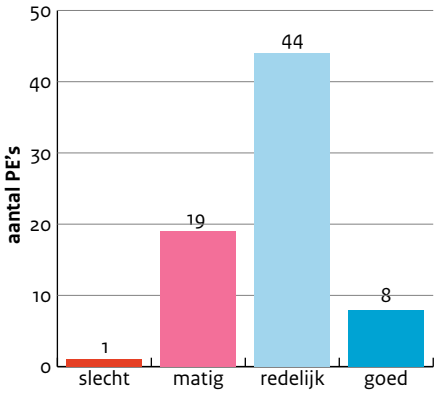
Met inachtneming van de onvermijdelijke beperking dat de onderzoekers door de organisaties zelf moesten worden geworven en betaald kan worden gesteld dat de organisaties zich serieus hebben ingespannen om gekwalificeerde, onafhankelijke externe deskundigen aan te trekken. En daarin ook meestal zijn geslaagd (Figuur 5–14).

Onderzoekverloop

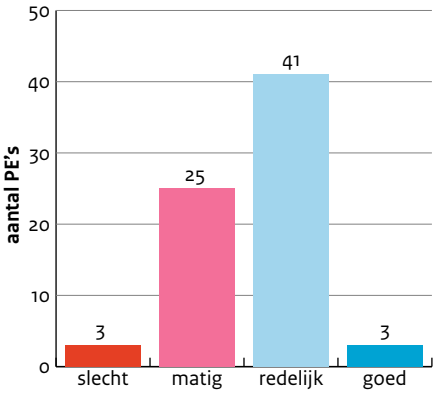
De meeste uitvoerders van de beoordeelde programma-evaluaties leggen op een behoorlijke manier en in voldoende detail verantwoording af over het verloop van hun onderzoek.

Uit Figuur 5–15 blijkt dat ruim twee derde van de PE op dit onderdeel “redelijk” tot “goed” scoorde.

Figuur 5–15 *Scores Verantwoording*



Figuur 5–16 *Scores Kwaliteitsbewaking*



Individuele PE’s werden in ruim een derde van de gevallen gesuperviseerd door directe betrokkenen die verantwoordelijk waren (geweest) voor de geëvalueerde programma’s; in een enkel geval was er helemaal geen begeleiding (Figuur 5–16). Daarentegen stond de uitvoering van de gezamenlijke PE’s zowel onder toezicht van een de Coördinatiegroep van het Partos Platform Kwaliteitshuis als van de Evaluatie Referentie Groep, die niet alleen deskundig bleek, maar ook geen blad voor de mond nam als het erom ging opdrachtgevers en evaluatoren op (dreigende) methodologische dwalingen te attenderen.

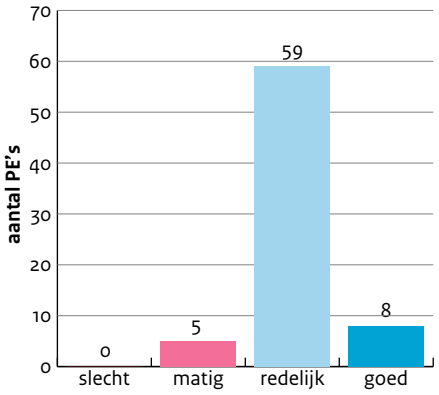
5.3.3 Bruikbaarheid

Voor de beoogde gebruikers lijkt de bruikbaarheid van onderzoekuitkomsten behalve van hun inhoud vaak vooral af te hangen van de meer pragmatische aspecten. De vier grafieken in deze paragraaf geven de scores op deze componenten weer.

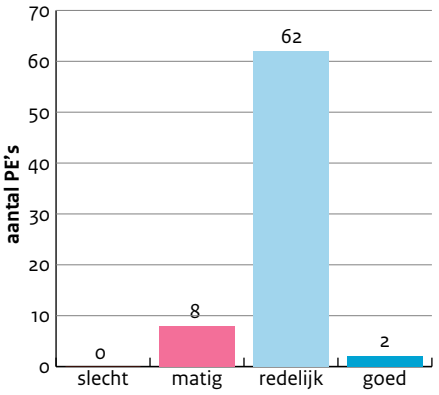
Presentatie

Bij de formulering van de onderzoekdoelstelling blijken zich vrijwel geen problemen te hebben voorgedaan (Figuur 5–17), al wordt het onderscheid met de probleemstelling niet altijd helder gemaakt.

Figuur 5–17 *Scores Onderzoekdoelstelling*



Figuur 5–18 *Scores Toegankelijkheid*

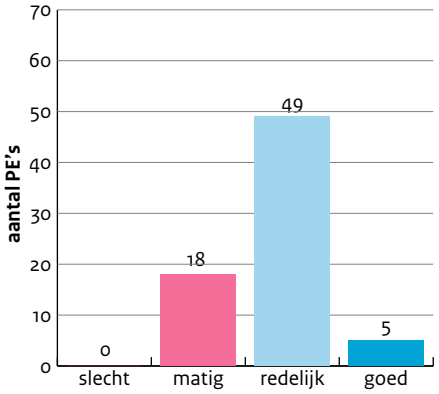


Ook op het punt van de toegankelijkheid van de onderzoekrapportages laten de beoordeelde PE's, zoals blijkt uit Figuur 5–18 weinig te wensen over.

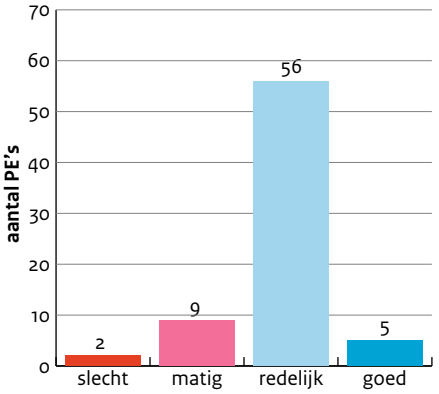
Aansluiting

In de meeste gevallen (Figuur 5–19) was de beantwoording van de onderzoekvragen redelijk tot goed te noemen, maar er kwamen toch ook opvallende missers voor. Dat laatste gebeurde vooral wanneer het onderzoekontwerp overbelast was geraakt door het al gesignaleerde probleem van de overdaad aan onderzoekvragen. In die gevallen werd een, soms aanzienlijk, deel van de vragen "gewoon" genegeerd, overigens zonder dat dit tot zichtbaar kritische reacties van opdrachtgevers leidde.

Figuur 5–19 *Scores Beantwoording Onderzoekvragen*



Figuur 5–20 *Scores Aanbevelingen*



In twee gevallen werden geen aanbevelingen gedaan. Zoals Figuur 5–20 toont, was de bruikbaarheid van de geformuleerde aanbevelingen bij de overige studies, op een beperkt aantal uitzonderingen na, ruim voldoende. Soms vergde de opvolging van aanbevelingen echter verstrekkende maatregelen of faciliteiten die zich niet onder de controle van de verantwoordelijke opdrachtgevers bevonden. In die gevallen kregen de aanbevelingen een vrijblijvend karakter.

5.4 Lerend vermogen

In § 2.2 werd gesteld dat het toezicht op de kwaliteit van programma-evaluaties (periodiek samengevat in meta-evaluaties zoals deze) behalve een verantwoordingsfunctie ook een leerfunctie kan vervullen. In hoeverre deze laatste uit de verf komt, hangt sterk af van het vermogen en de bereidheid tot leren bij de betrokken organisaties. Hieraan wordt door het DGIS dan ook – terecht – veel waarde gehecht. Zoveel dat het één van de criteria vormde waarop subsidie-aanvragers vóór hun toelating tot het MFS werden beoordeeld. Daarbij was de inschatting van de evaluatiefunctie het belangrijkste ingrediënt voor de oordeelsvorming over het lerend vermogen.

5.4.1 Evaluatiefunctiebeoordeling – ex ante vs. ex post

De aanvragen die 114 particuliere organisaties in 2006 indienden voor de toekenning van een subsidie onder het MFS-I werden beoordeeld door de onafhankelijke Commissie Bikker. Deze hanteerde hiervoor een beoordelingskader dat 33 criteria omvatte, verdeeld over drie afzonderlijke toetsen, te weten een drempeltoets, een organisatietoets (O-toets) en een voorsteltoets (V-toets).[35] Van dit drietal was de organisatietoets specifiek gericht op de (ex ante) beoordeling van de kwaliteit van de bedrijfsvoering van de aanvragende organisatie.

Hiertoe telde deze O-toets tien criteria waarvan er één, aangeduid als het *Lerend vermogen in de keten*, specifiek gericht was op de inrichting en werking van de evaluatiefunctie van de indiener. Volgens het aanvraagstramien dat de aanvragers verplicht waren te hanteren, werd dit criterium geoperationaliseerd in verschillende vragen. Deze hadden betrekking op de verankering van evaluatie in het beleid, het toezicht op de evaluatiekwaliteit en de terugkoppeling van evaluatieresultaten naar beleidsaanpassing en –vernieuwing. Van de 100 punten die met de tien criteria van de O-toets in totaal konden worden gescoord, waren er maximaal negen te verdienen met de toetsing aan het criterium *Lerend vermogen in de keten*.

[35] Opzet en verloop van het beoordelingsproces zijn uitvoerig beschreven in het rapport van de commissie *Kwaliteit in ontwikkeling*.

De elf MFS-organisaties wier programma-evaluaties het object van dit onderzoek vormen, behaalden met driekwart tot 100% van het maximale aantal punten hoge tot zeer hoge scores:

Tabel 5–2 Scores Organisatietoets: *Lerend vermogen*

MFS-organisatie	Score (maximum 9)
Cordaid	9
Hivos	9
ICCO	8
Oxfam Novib	8
Terre des Hommes	7
ETC	8
ICS	7
IKV/Pax Christi	8
IUCN	8
Solidaridad	8
Woord en Daad	9
Gemiddeld	**8**

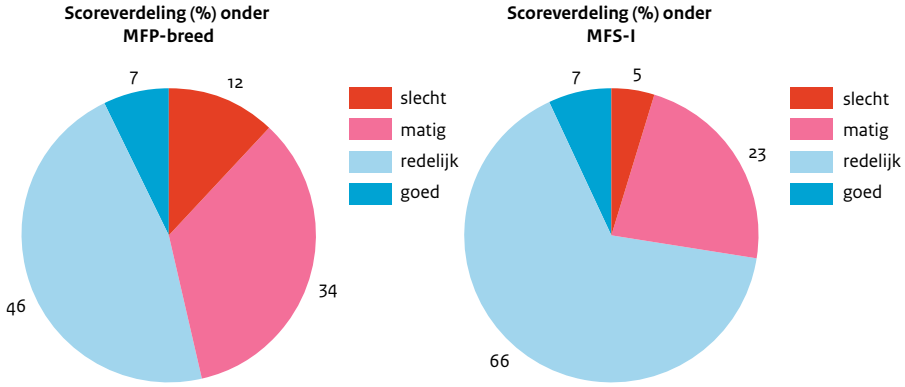
De gemiddelde score van de elf organisaties lag op 89% van het maximum. Daarmee nam deze score tevens de derde plaats in onder de gemiddelde scores op de tien beoordelingscriteria die samen de O-toets vormden. Kortom, de waardering voor de manier waarop de aanvragende organisaties hun evaluatiefuncties presenteerden, was niet alleen in absolute, maar ook in relatieve zin hoog. Dit roept de vraag op in hoeverre deze (zeer) positieve *ex ante* beoordeling van de evaluatiefuncties van de organisaties strookt met de *ex post* vastgestelde kwaliteit van de door hen ondernomen programma-evaluaties. Om hier een indruk van te geven zijn op de scorekaarten voor de elf organisaties in bijlage 5 de bovenstaande O-toets scores overgenomen. Op deze kaarten zijn de beoordeelde PE's genummerd in de volgorde waarin ze werden uitgevoerd. Hoewel de PE-scores niet zijn geaggregeerd (zie § 4.1.3), kost het weinig moeite het contrast op te merken tussen enerzijds de hoogte van de in 2006 *ex ante* toegekende waardering voor de evaluatiefuncties en anderzijds de talrijke tekortkomingen die de *ex post* beoordeelde PE's, zeker in het begin van de MFS-I subsidieperiode vertoonden. De empirische basis van de vaststelling van de PE kwaliteiten leidt dan ook tot de plaatsing van vraagtekens bij de voorspellende waarde van de *ex ante* beoordeling van de evaluatiefunctie als onderdeel van de organisatietoets.

5.4.2 Leereffecten

Of de leerfunctie van het toezicht op de kwaliteit van de programma-evaluaties inhoud krijgt, hangt, zoals eerder werd betoogd, vooral af van de betreffende organisatie. De scorekaarten in bijlage 5 geven een indruk van de ontwikkelingen op het gebied van evaluatiekwaliteit in de loop van de tijd. Hierbij zijn drie (sub)groepen zijn te onderscheiden.

1. Organisaties waarvan zowel onder het MFP-breed als onder het MFS-I de programma-evaluaties werden beoordeeld. Dit zijn de oorspronkelijke vier MFO's (Cordaid, Hivos, ICCO en Oxfam Novib) plus hun samenwerkingsverband (eerst MBN, later Partos).Voor deze organisaties kan de scoreverdeling uit de MFP-breed periode (2003 t/m/ 2006) worden vergeleken met die uit de MFS-I periode (2007 t/m 2010). Beide verdelingen zijn weergegeven in Figuur 5–21, die een onmiskenbare kwaliteitsontwikkeling in positieve richting laat zien.

Figuur 5–21 *Scorevergelijking MFP-breed – MFS-I Programma-Evaluaties van Cordaid, Hivos, Icco en Oxfam Novib*



Deze ontwikkeling betreft zowel de individuele PE's als de gezamenlijke onderzoeken in MBN- of Partos-verband.

2. De overige zeven 'grote' subsidieontvangers kwamen voor het eerst onder het MFS-I met het fenomeen PE-beoordeling in aanraking en voor hen is de periode waarin kwaliteits-veranderingen konden worden vastgesteld dan ook navenant korter. Toch zijn ook hier twee subgroepen te onderscheiden.

   a. Organisaties die hun programma-evaluaties min of meer *gebundeld* en meestal tegen het einde van de subsidieperiode ter beoordeling voorlegden, bijvoorbeeld omdat ze pas in dat stadium genoeg evalueerbare uitgaven hadden gedaan om een voldoende groot volume aan bestedingen te kunnen bestrijken. Bij deze gebun-

delde aanlevering werden geen systematische verschillen tussen de kwaliteitsbeoordelingen van de individuele studies zichtbaar Tot deze (sub)groep behoren International Child Support, IKV/Pax Christi en Woord & Daad.

b. Organisaties die hun programma-evaluatierapporten *gespreid* over de subsidieperiode ter beoordeling aanleverden, zodat zij bij later uitgevoerde onderzoeken hun voordeel konden doen met de commentaren en suggesties die ze hadden ontvangen op de validiteit, betrouwbaarheid en bruikbaarheid van eerder beoordeelde studies. ETC, Solidaridad en Terre des Hommes hanteerden deze aanpak. Hoewel het per organisatie uiteraard slechts om enkele PE's ging (respectievelijk twee, drie en vier studies), toonden sommige evaluatievertegenwoordigers zich zo bedreven in de absorptie en verwerking van kritiek en adviezen, dat zelfs binnen de geringe aantallen onderzoeken positieve kwaliteitsontwikkelingen zichtbaar lijken te worden. Deze wijzen in de richting van een optredend leereffect. De omvang van het beschikbare materiaal is echter te gering om op dit punt verdergaande conclusies te trekken.

Met slechts één, zij het veelomvattende, programma-evaluatie valt IUCN uiteraard buiten deze indeling.

Bijlagen

Bijlage 1 Over IOB

Doelstellingen

De Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (IOB) stelt zich ten doel een bijdrage te leveren aan de kennis over de uitvoering en effecten van het Nederlands buitenlands beleid. IOB voorziet in de behoefte aan onafhankelijke evaluatie van beleid en uitvoering ten aanzien van alle beleidsterreinen die vallen binnen de homogene groep buitenlanduitgaven (HGIS). Voorts adviseert IOB ten aanzien van de programmering en uitvoering van de evaluaties die onder verantwoordelijkheid van beleidsdirecties en ambassades worden gedaan.

De evaluaties stellen de bewindspersonen in staat om aan het parlement *verantwoording* af te leggen over het gevoerde beleid en de besteding van middelen. Bij de uitvoering van evaluaties wordt naast verantwoording ook aandacht geschonken aan *leren*. Daarbij wordt gestreefd naar inpassing van de resultaten van de evaluatieonderzoeken in de beleidscyclus van het ministerie van Buitenlandse Zaken. De rapporten die uit het onderzoek voortvloeien worden gebruikt als gerichte feedback om zowel beleidsvorming als uitvoering te verbeteren. Gewapend met de kennis over de resultaten van het gevoerde beleid kunnen beleidsmakers nieuwe interventies beter en doelgerichter voorbereiden.

Organisatie en Kwaliteitsborging

IOB beschikt over een staf van ervaren onderzoekers en een eigen budget. Bij de uitvoering van evaluaties maakt IOB ook gebruik van externe deskundigen met specialistische kennis van het onderwerp van onderzoek. Ten behoeve van de kwaliteitsbewaking stelt IOB voorts voor elke evaluatie een referentiegroep samen waarin naast externe deskundigen ook belanghebbenden binnen het ministerie zitting hebben. Het Panel van Advies heeft als taak om het gebruik en de bruikbaarheid van evaluatiestudies te bevorderen middels gevraagde en ongevraagde adviezen. Dit Panel bestaat uit vier gerenommeerde externe deskundigen en de adviezen zijn onafhankelijk en openbaar.

Programmering

De IOB-evaluatieprogrammering maakt deel uit van het overzicht van voorgenomen evaluaties dat, gerangschikt naar beleidsartikel, is opgenomen in de Memorie van Toelichting van de Begroting van Buitenlandse Zaken. IOB heeft eindverantwoordelijkheid voor de evaluatieprogrammering op het terrein van ontwikkelingssamenwerking, en adviseert over de programmering op het terrein van buitenlandbeleid. De keuze van onderwerpen wordt bepaald door de behoefte vanuit het parlement, vragen vanuit het departement en ontwikkelingen in de samenleving.

Aanpak en methodologie

Oorspronkelijk stonden de activiteiten van IOB in het teken van aparte projectevaluaties ten behoeve van de minister voor Ontwikkelingssamenwerking. Na 1985 werden de onderzoeken omvangrijker en richtten zich op sectoren, thema's of landen. Bovendien werden de rapporten van IOB aan het parlement aangeboden en daarmede openbaar. In 1996 werd – in het kader van de herijking van het buitenlands beleid en de reorganisatie van het ministerie

van Buitenlandse Zaken – het werkterrein van IOB uitgebreid tot het volledige buitenlandse beleid van de Nederlandse overheid. De naam van de Inspectie werd gewijzigd van IOV (Inspectie Ontwikkelingssamenwerking te Velde) in IOB. Voorts wordt vanaf de jaren negentig nauw samengewerkt met evaluatiediensten van andere landen, onder meer in de uitvoering van gezamenlijke evaluaties en binnen het OECD-DAC evaluatienetwerk.

IOB streeft op methodologisch gebied naar hoogwaardige kwaliteit en vernieuwing. Het evaluatieonderzoek maakt bij voorkeur gebruik van een interactieve combinatie van kwantitatieve en kwalitatieve onderzoeksmethoden. Op verschillende beleidsterreinen worden thans robuuste impactstudies uitgevoerd. Voorts worden systematische overzichtsstudies verricht van beschikbare empirische resultaten op prioritaire beleidsterreinen.

Bijlage 2 Algemene kenmerken Programma-evaluaties

Cordaid	**Common Programming and Shared Indicators**				
locatie	Vietnam		evaluatieperiode	2004 – 2006	
financieel belang	€ 800.000	evaluatie-uitgaven	€19.484	uitgaven/ belang	2,4 %
doorlooptijd	4 mndn	rapportdatum	apr 2007	omvang	79 p
Cordaid	**PHC Support Programme in Cyangugu Province**				
locatie	Rwanda		evaluatieperiode	2003 – 2005	
financieel belang	€ 2.500.000	evaluatie-uitgaven	€ 22.588	uitgaven/ belang	0,9 %
doorlooptijd	4 mndn	rapportdatum	apr 2007	omvang	66 p
Cordaid	**Biologische Fruitteelt Latijns-Amerika**				
locatie	Bolivia, Colombia, Dominicaanse Republiek, Peru		evaluatieperiode	2004 – 2007	
financieel belang	€ 2.300.000	evaluatie-uitgaven	€ 55.000	uitgaven/ belang	2,4 %
doorlooptijd	11 mndn	rapportdatum	okt 2007	omvang	347 p
Cordaid	**Kinderprogramma**				
locatie	Peru, Bolivia, India, Filippijnen		evaluatieperiode	2002 – 2007	
financieel belang	€ 4.600.000	evaluatie-uitgaven	€ 76.000	uitgaven/ belang	1,7 %
doorlooptijd	19 mndn	rapportdatum	jul 2008	omvang	224 p
Cordaid	**Midden-Oosten en Midden- en Oost-Europa Programma (MOMOE)**				
locatie	Georgië, Moldavië, Albanië		evaluatieperiode	2003 – 2006	
financieel belang	€ 9.667.000	evaluatie-uitgaven	€ 103.975	uitgaven/ belang	1,1 %
doorlooptijd	15 mndn	rapportdatum	aug 2008	omvang	207 p
Cordaid	**Capaciteitsversterking Financieel Management**				
locatie	India, Peru, Uganda		evaluatieperiode	2003 – 2006	
financieel belang	€ 2.600.000	evaluatie-uitgaven	€ 109.000	uitgaven/ belang	4,2 %
doorlooptijd	30 mndn	rapportdatum	jan 2009	omvang	247 p
Cordaid	**Sloppenwijkbewoners**				
locatie	Mali, Peru, Zuid-Afrika		evaluatieperiode	2003 – 2007	
financieel belang	€ 7.241.037	evaluatie-uitgaven	€ 99.804	uitgaven/ belang	1,4 %
doorlooptijd	13 mndn	rapportdatum	mei 2009	omvang	577 p

Cordaid	Performance-Based Financing				
locatie	Congo, Burundi, Tanzania, Zambia		evaluatieperiode	2006 – 2008	
financieel belang	€ 3.804.556	evaluatie-uitgaven	€ 179.749	uitgaven/ belang	4,7 %
doorlooptijd	12 mndn	rapportdatum	jun 2009	omvang	505 p
Cordaid	**SME Support and Value Chain Development**				
locatie	wereldwijd		evaluatieperiode	2002 – 2008	
financieel belang	€ 18.000.000	evaluatie-uitgaven	€ 107.125	uitgaven/ belang	0,6 %
doorlooptijd	14 mndn	rapportdatum	dec 2009	omvang	755 p
Cordaid	**Disaster Risk Reduction**				
locatie	wereldwijd		evaluatieperiode	2004 – 2008	
financieel belang	€ 12.830.713	evaluatie-uitgaven	€ 97.000	uitgaven/ belang	0,8 %
doorlooptijd	13 mndn	rapportdatum	feb 2010	omvang	195 p
Cordaid	**HIV Programma**				
locatie	wereldwijd		evaluatieperiode	2005 – 2009	
financieel belang	€ 19.990.000	evaluatie-uitgaven	€ 201.500	uitgaven/ belang	1,0 %
doorlooptijd	13 mndn	rapportdatum	mrt 2011	omvang	277 p
Hivos	**LGBT Emancipation in Southern Africa**				
locatie	Zuidelijk Afrika		evaluatieperiode	1995 – 2006	
financieel belang	€ 2.842.000	evaluatie-uitgaven	€ 77.672	uitgaven/ belang	2,7 %
doorlooptijd	23 mndn	rapportdatum	okt 2007	omvang	309 p
Hivos	**Missing Links: Growing Organic Chains between Farmer and Market**				
locatie	Oost-Afrika		evaluatieperiode	2000 – 2006	
financieel belang	€ 6.900.000	evaluatie-uitgaven	€ 100.000	uitgaven/ belang	1,4 %
doorlooptijd	13 mndn	rapportdatum	mei 2008	omvang	89 p
Hivos	**Evaluación del Programa Arte y Cultura**				
locatie	Midden-Amerika		evaluatieperiode	2002 – 2006	
financieel belang	€ 3.200.000	evaluatie-uitgaven	€ 100.000	uitgaven/ belang	3,1 %
doorlooptijd	11 mndn	rapportdatum	mei 2008	omvang	251 p
Hivos	**Tsunami Response**				
locatie	Indonesië, India, Sri Lanka		evaluatieperiode	2004 – 2007	
financieel belang	€ 2.878.221	evaluatie-uitgaven	€ 24.942	uitgaven/ belang	0,9 %
doorlooptijd	10 mndn	rapportdatum	jul 2008	omvang	83 p

Hivos	Human Rights and Democratisation in Kenya				
locatie	Kenya		evaluatieperiode	1996 – 2007	
financieel belang	€ 4.300.000	evaluatie-uitgaven	€ 70.723	uitgaven/ belang	1,6 %
doorlooptijd	13 mndn	rapportdatum	dec 2008	omvang	331 p
Hivos	**ICT en Media Programme**				
locatie	wereldwijd		evaluatieperiode	2005 – 2008	
financieel belang	€ 13.542.218	evaluatie-uitgaven	€ 120.385	uitgaven/ belang	0,9 %
doorlooptijd	15 mndn	rapportdatum	nov 2009	omvang	493 p
Hivos	**HIV & AIDS Programme**				
locatie	India, Zuid-Amerika		evaluatieperiode	2000 – 2009	
financieel belang	€ 6.900.000	evaluatie-uitgaven	€ 89.250	uitgaven/ belang	1,3 %
doorlooptijd	12 mndn	rapportdatum	jan 2011	omvang	176 p
ICCO	**Stille Oceaan Programma**				
locatie	Papoea-Nieuw-Guinea, Salomonseilanden, Fiji		evaluatieperiode	2002 – 2005	
financieel belang	€ 4.320.000	evaluatie-uitgaven	€ 61.410	uitgaven/ belang	1,4 %
doorlooptijd	11 mndn	rapportdatum	mrt 2006	omvang	76 p
ICCO	**Kameroen Programma**				
locatie	Kameroen		evaluatieperiode	2002 – 2006	
financieel belang	€ 3.160.340	evaluatie-uitgaven	€ 48.736	uitgaven/ belang	1,5 %
doorlooptijd	7 mndn	rapportdatum	sep 2006	omvang	73 p
ICCO	**ACT Lobbycampagne**				
locatie	Indonesië		evaluatieperiode	2001 – 2005	
financieel belang	€ 52.000	evaluatie-uitgaven	€ 10.910	uitgaven/ belang	21 %
doorlooptijd	5 mndn	rapportdatum	nov 2006	omvang	46 p
ICCO	**Strong People, Strong State**				
locatie	Vietnam		evaluatieperiode	2002 – 2006	
financieel belang	€ 7.807.527	evaluatie-uitgaven	€ 25.510	uitgaven/ belang	0,3 %
doorlooptijd	8 mndn	rapportdatum	apr 2008	omvang	67 p
ICCO	**Technical and Vocational Education and Training**				
locatie	Albanië, Ethiopië, India, Kenya		evaluatieperiode	2003 – 2006	
financieel belang	€ 18.408.768	evaluatie-uitgaven	€ 118.124	uitgaven/ belang	0,6 %
doorlooptijd	12 mndn	rapportdatum	apr 2008	omvang	431 p

ICCO	Terrafina Microfinance				
locatie	Burkina Faso, Burundi, Congo, Ethiopië, Mali, Rwanda, Senegal		evaluatieperiode	2005 – 2007	
financieel belang	€ 5.360.000	evaluatie-uitgaven	€ 23.850	uitgaven/ belang	0,4 %
doorlooptijd	5 mndn	rapportdatum	feb 2009	omvang	74 p
ICCO	**International Markets Programme**				
locatie	wereldwijd		evaluatieperiode	2005 – 2009	
financieel belang	€ 37.613.605	evaluatie-uitgaven	€ 141.502	uitgaven/ belang	0,4 %
doorlooptijd	15 mndn	rapportdatum	nov 2010	omvang	460 p
ICCO	**Local Market Development Programme**				
locatie	wereldwijd		evaluatieperiode	2007 – 2010	
financieel belang	€ 41.124.961	evaluatie-uitgaven	€ 162.305	uitgaven/ belang	0,4 %
doorlooptijd	7 mndn	rapportdatum	dec 2010	omvang	324 p
ICCO	**Food Security Programme**				
locatie	wereldwijd		evaluatieperiode	2007 – 2010	
financieel belang	€ 28.603.068	evaluatie-uitgaven	€ 167.511	uitgaven/ belang	0,6 %
doorlooptijd	9 mndn	rapportdatum	jan 2011	omvang	798 p
Oxfam Novib	**Core Country Programme Evaluation Nigeria**				
locatie	Nigeria		evaluatieperiode	1997 – 2006	
financieel belang	€ 4.909.000	evaluatie-uitgaven	€ 120.000	uitgaven/ belang	2,4 %
doorlooptijd	19 mndn	rapportdatum	mrt 2007	omvang	106 p
Oxfam Novib	**Core Country Programme Evaluation Somalia**				
locatie	Somalia		evaluatieperiode	2001 – 2006	
financieel belang	€ 8.798.000	evaluatie-uitgaven	€ 170.000	uitgaven/ belang	1,9 %
doorlooptijd	20 mndn	rapportdatum	aug 2007	omvang	146 p
Oxfam Novib	**Core Country Programme Evaluation India**				
locatie	India		evaluatieperiode	2002 – 2006	
financieel belang	€ 17.553.000	evaluatie-uitgaven	€ 150.000	uitgaven/ belang	0,9 %
doorlooptijd	13 mndn	rapportdatum	mei 2008	omvang	95 p
Oxfam Novib	**Review Linkis Oxfam Novib –Draagvlakversterking in Nederland**				
locatie	wereldwijd		evaluatieperiode	2004 – 2009	
financieel belang	€ 14.200.000	evaluatie-uitgaven	€ 36.295	uitgaven/ belang	0,3 %
doorlooptijd	2 mndn	rapportdatum	jun 2009	omvang	78 p

Oxfam Novib	**Core Country Programme Evaluation Angola**				
locatie	Angola		evaluatieperiode	2002 – 2006	
financieel belang	€ 5.835157	evaluatie-uitgaven	€ 209.897	uitgaven/ belang	3,6 %
doorlooptijd	25 mndn	rapportdatum	jun 2009	omvang	283 p
Oxfam Novib	**Humanitarian Capacity Building**				
locatie	wereldwijd		evaluatieperiode	2005 – 2008	
financieel belang	€ 5.751.142	evaluatie-uitgaven	€ 0	uitgaven/ belang	0 %
doorlooptijd	7 mndn	rapportdatum	jul 2009	omvang	67 p
Oxfam Novib	**Core Country Programme Evaluation Palestinians in the Middle East**				
locatie	Palestijnse Gebieden		evaluatieperiode	2003 – 2007	
financieel belang	€ 7.827.000	evaluatie-uitgaven	€ 148.000	uitgaven/ belang	1,9 %
doorlooptijd	12 mndn	rapportdatum	okt 2009	omvang	153 p
Oxfam Novib	**Moving a Mountain – Stop EPA Campaign**				
locatie	wereldwijd		evaluatieperiode	2004 – 2008	
financieel belang	€ 884.000	evaluatie-uitgaven	€ 40.140	uitgaven/ belang	4,5 %
doorlooptijd	25 mndn	rapportdatum	jan 2010	omvang	118 p
Oxfam Novib	**Core Country Programme Evaluation Ethiopia**				
locatie	Ethiopië		evaluatieperiode	2003 – 2009	
financieel belang	€ 16.426.893	evaluatie-uitgaven	€ 168.000	uitgaven/ belang	1,0 %
doorlooptijd	13 mndn	rapportdatum	jan 2010	omvang	222 p
Oxfam Novib	**Global Programme**				
locatie	wereldwijd		evaluatieperiode	2005 – 2008	
financieel belang	€ 18.152.000	evaluatie-uitgaven	€ 229.159	uitgaven/ belang	1,0 %
doorlooptijd	11 mndn	rapportdatum	mrt 2010	omvang	704 p
Oxfam Novib	**Core Country Programme Evaluation Burundi**				
locatie	Burundi		evaluatieperiode	2005 – 2009	
financieel belang	€ 5.023.783	evaluatie-uitgaven	€ 89.880	uitgaven/ belang	1,8 %
doorlooptijd	8 mndn	rapportdatum	dec 2010	omvang	209 p
Oxfam Novib	**Evaluación del Programa de América Latina**				
locatie	Latijns-America		evaluatieperiode	2003 – 2009	
financieel belang	€ 56.866.000	evaluatie-uitgaven	€ 190.000	uitgaven/ belang	0,3 %
doorlooptijd	13 mndn	rapportdatum	feb 2011	omvang	323 p

ETC	PROLINNOVA				
locatie	Cambodja, Ethiopië, Ghana, Nepal, Niger, Zuid-Afrika, Sudan, Tanzania, Uganda		evaluatieperiode	2004 – 2008	
financieel belang	€ 2.017.000	evaluatie-uitgaven	€ 43.003	uitgaven/ belang	2,1 %
doorlooptijd	7 mndn	rapportdatum	mrt 2009	omvang	155 p
ETC	**Energy Access Programme**				
locatie	wereldwijd		evaluatieperiode	2007 – 2010	
financieel belang	€ 3.1354.724	evaluatie-uitgaven	€ 44.990	uitgaven/ belang	1,4 %
doorlooptijd	11 mndn	rapportdatum	nov 2010	omvang	195 p
ICS	**Rice Bank Evaluation**				
locatie	Cambodja, Laos, Thailand		evaluatieperiode	2004 – 2008	
financieel belang	€ 322.689	evaluatie-uitgaven	€ 5.713	uitgaven/ belang	1,7 %
doorlooptijd	5 mndn	rapportdatum	aug 2008	omvang	56 p
ICS	**Action for Children Program**				
locatie	Brazilië, India, Zuid-Afrika		evaluatieperiode	2007 – 2010	
financieel belang	€ 2.285.000	evaluatie-uitgaven	€ 62.775	uitgaven/ belang	2.7 %
doorlooptijd	4 mndn	rapportdatum	jun 2010	omvang	62p
ICS	**Aflatoun Secretariat**				
locatie	wereldwijd		evaluatieperiode	2007 – 2010	
financieel belang	€ 2.158.000	evaluatie-uitgaven	€ 10.500	uitgaven/ belang	0,5 %
doorlooptijd	5 mndn	rapportdatum	jun 2010	omvang	41 p
ICS	**Knowledge Centre on Civic Driven Child Development (CDCD)**				
locatie	wereldwijd		evaluatieperiode	2007 – 2009	
financieel belang	€ 680.745	evaluatie-uitgaven	€ 31.167	uitgaven/ belang	4,6 %
doorlooptijd	6 mndn	rapportdatum	jun 2010	omvang	114 p
ICS	**African Child Policy Forum (ACPF)**				
locatie	Afrika		evaluatieperiode	2007 – 2009	
financieel belang	€ 1.947.685	evaluatie-uitgaven	€ 16.748	uitgaven/ belang	0,9 %
doorlooptijd	2 mndn	rapportdatum	jul 2010	omvang	70 p

ICS	CDCD Programme Cambodia				
locatie	Cambodja		evaluatieperiode	2008 – 2009	
financieel belang	€ 613.004	evaluatie-uitgaven	€ 21.970	uitgaven/ belang	3,6 %
doorlooptijd	4 mndn	rapportdatum	jul 2010	omvang	43 p
ICS	**Change for Children (CfC)**				
locatie	Suriname		evaluatieperiode	2007 – 2010	
financieel belang	€ 2.038.749	evaluatie-uitgaven	€ 24.100	uitgaven/ belang	1,2 %
doorlooptijd	8 mndn	rapportdatum	okt 2010	omvang	86 p
ICS	**Child Rights in School Improvement Programme**				
locatie	Kenya		evaluatieperiode	2007 – 2010	
financieel belang	€ 1.415.604	evaluatie-uitgaven	€ 34.904	uitgaven/ belang	2,5 %
doorlooptijd	4 mndn	rapportdatum	nov 2010	omvang	73 p
IKV/Pax Christi	**Local Peace Initiatives in Ituri, DRC**				
locatie	Congo		evaluatieperiode	2003 – 2006	
financieel belang	€ 288.482	evaluatie-uitgaven	€ 32.170	uitgaven/ belang	11 %
doorlooptijd	8 mndn	rapportdatum	okt 2007	omvang	60 p
IKV/Pax Christi	**The European Coalition on Oil in Sudan**				
locatie	Sudan		evaluatieperiode	2001 – 2007	
financieel belang	€ 451.661	evaluatie-uitgaven	€ 15.040	uitgaven/ belang	3,3 %
doorlooptijd	22 mndn	rapportdatum	jul 2008	omvang	55 p
IKV/Pax Christi	**Giving the Syrian Voices of Change More Than Some Extra Oxygen!**				
locatie	Syrië		evaluatieperiode	2003 – 2009	
financieel belang	€ 334.993	evaluatie-uitgaven	€ 8.698	uitgaven/ belang	2,5 %
doorlooptijd	2 mndn	rapportdatum	nov 2009	omvang	42 p
IKV/Pax Christi	**The Cluster Munition Coalition**				
locatie	wereldwijd		evaluatieperiode	2003 – 2008	
financieel belang	€ 60.045	evaluatie-uitgaven	€ 11.573	uitgaven/ belang	19 %
doorlooptijd	2 mndn	rapportdatum	nov 2009	omvang	57 p
IKV/Pax Christi	**United Civilians for Peace (UCP)**				
locatie	Nederland, Palestijnse Gebieden, Israël		evaluatieperiode	2007 – 2009	
financieel belang	€ 125.805	evaluatie-uitgaven	€ 945	uitgaven/ belang	0,8 %
doorlooptijd	3 mndn	rapportdatum	apr 2010	omvang	45 p

IKV/Pax Christi	Kosovo Programma				
locatie	Kosovo		evaluatieperiode	2007 – 2009	
financieel belang	€ 771.319	evaluatie-uitgaven	€ 7.250	uitgaven/ belang	0,9 %
doorlooptijd	9 mndn	rapportdatum	mei 2010	omvang	33 p
IKV/Pax Christi	**Nagorno Karabach Programma**				
locatie	Armenië, Azerbeidzjan		evaluatieperiode	2009 – 2010	
financieel belang	€ 233.788	evaluatie-uitgaven	€ 14.757	uitgaven/ belang	6,3 %
doorlooptijd	6 mndn	rapportdatum	jun 2010	omvang	45 p
IKV/Pax Christi	**Sudan Integrated Peace building Programme (SIPP)**				
locatie	Sudan		evaluatieperiode	2007 – 2010	
financieel belang	€ 428.783	evaluatie-uitgaven	€ 17.400	uitgaven/ belang	4,1 %
doorlooptijd	5 mndn	rapportdatum	nov 2010	omvang	47 p
IKV/Pax Christi	**Norte del Cauca Programa**				
locatie	Colombia		evaluatieperiode	2007 – 2010	
financieel belang	€ 409.000	evaluatie-uitgaven	€ 9.990	uitgaven/ belang	2,4 %
doorlooptijd	7 mndn	rapportdatum	mrt 2011	omvang	56 p
IUCN	**Ecosystems and Human Well-Being Programme**				
locatie	wereldwijd		evaluatieperiode	2007 – 2009	
financieel belang	€ 18.337.628	evaluatie-uitgaven	€ 54.246	uitgaven/ belang	0,3 %
doorlooptijd	10 mndn	rapportdatum	dec 2009	omvang	114 p
Solidaridad	**The Impact of Fair Trade**				
locatie	Peru, Costa Rica, Ghana		evaluatieperiode	2002 – 2008	
financieel belang	€ 1.216.875	evaluatie-uitgaven	€ 83.117	uitgaven/ belang	6.8 %
doorlooptijd	17 mndn	rapportdatum	aug 2008	omvang	95 p
Solidaridad	**Food, Feed and Fuel for Thought**				
locatie	wereldwijd		evaluatieperiode	2007 – 2010	
financieel belang	€ 2.262.964	evaluatie-uitgaven	€ 50.000	uitgaven/ belang	2,2 %
doorlooptijd	9 mndn	rapportdatum	apr 2010	omvang	155 p
Solidaridad	**Evaluation Fashion Programme Solidaridad 2003 – 2010**				
locatie	wereldwijd		evaluatieperiode	2007 – 2010	
financieel belang	€ 3.584.180	evaluatie-uitgaven	€ 59.024	uitgaven/ belang	1,6 %
doorlooptijd	5 mndn	rapportdatum	dec 2010	omvang	345 p

Terre des Hommes	**Pathuakali Development Project (PDP) & Coastal Island Health and Education Project (CIHEP)**				
locatie	Bangladesh		evaluatieperiode	2003 – 2008	
financieel belang	€ 2.193.216	evaluatie-uitgaven	€ 33.489	uitgaven/ belang	1,5 %
doorlooptijd	8 mndn	rapportdatum	feb 2009	omvang	62 p
Terre des Hommes	**Mathagondapalli Education Centre (MEC)**				
locatie	India		evaluatieperiode	2000 – 2008	
financieel belang	€ 422.738	evaluatie-uitgaven	€ 1.043	uitgaven/ belang	0,2 %
doorlooptijd	8 mndn	rapportdatum	apr 2009	omvang	103 p
Terre des Hommes	**Southeast Asia Campaign against Child Trafficking**				
locatie	Zuid-Oost Azië		evaluatieperiode	2006 – 2010	
financieel belang	€ 3.175.130	evaluatie-uitgaven	€ 30.000	uitgaven/ belang	0,9 %
doorlooptijd	7 mndn	rapportdatum	jan 2011	omvang	80 p
Terre des Hommes	**Onderwijs-en Gezondheidsprogramma's**				
locatie	wereldwijd		evaluatieperiode	2007 – 2010	
financieel belang	€ 19.724.592	evaluatie-uitgaven	€ 163.120	uitgaven/ belang	0,8 %
doorlooptijd	7 mndn	rapportdatum	mrt 2011	omvang	353 p
Woord en Daad	**Education Programme**				
locatie	wereldwijd		evaluatieperiode	2004 – 2007	
financieel belang	€ 1.495.378	evaluatie-uitgaven	€ 2.078	uitgaven/ belang	0,1 %
doorlooptijd	17 mndn	rapportdatum	feb 2009	omvang	217 p
Woord en Daad	**Job and Income Programme**				
locatie	wereldwijd		evaluatieperiode	2005 – 2008	
financieel belang	€ 9.694.881	evaluatie-uitgaven	€ 41.051	uitgaven/ belang	0,4 %
doorlooptijd	8 mndn	rapportdatum	mei 2009	omvang	148 p
Woord en Daad	**Basic Needs Programme**				
locatie	wereldwijd		evaluatieperiode	2005 – 2008	
financieel belang	€ 7.172.578	evaluatie-uitgaven	€ 41.051	uitgaven/ belang	0,6 %
doorlooptijd	13 mndn	rapportdatum	mei 2009	omvang	107 p

Partos	Conflict Transformation: a Science and an Art				
locatie	Grote Merengebied, Indonesië, Sierra Leone		evaluatieperiode	2003 – 2006	
financieel belang	€ 29.100.000	evaluatie-uitgaven	€ 376.223	uitgaven/ belang	1,3 %
doorlooptijd	16 mndn	rapportdatum	nov 2008	omvang	661 p
Partos	**Microfinanciering**				
locatie	wereldwijd		evaluatieperiode	2003 – 2007	
financieel belang	€ 172.300.215	evaluatie-uitgaven	€ 454.852	uitgaven/ belang	0,3 %
doorlooptijd	21 mndn	rapportdatum	feb 2010	omvang	944 p
Partos	**Inheemse Volken**				
locatie	Ethiopië, Kenya, Bolivia, Guatemala, India		evaluatieperiode	2003 – 2008	
financieel belang	€ 32.447.595	evaluatie-uitgaven	€ 338.227	uitgaven/ belang	1,0 %
doorlooptijd	30 mndn	rapportdatum	jul 2010	omvang	532 p

Bijlage 3 Beoordelingslijst Programma-evaluaties

Deze beoordelingslijst wordt gehanteerd om de opzet en uitvoering van Programma-evaluaties (PE's)[36] binnen het Medefinancieringsstelsel (MFS) te toetsen aan gangbare kwaliteitseisen voor sociaalwetenschappelijk onderzoek.
Daartoe wordt eerst een aantal basiskenmerken van de betreffende PE geregistreerd. Deze kenmerken hebben achtereenvolgens betrekking op het geëvalueerde object, het uitgevoerde onderzoek, de daarbij betrokken actoren en het onderzoeksrapport. Vervolgens vindt de toetsing plaats aan de drie kwaliteitscriteria: Validiteit, Betrouwbaarheid en Bruikbaarheid. Deze criteria zijn geoperationaliseerd door de definiëring van een aantal indicatoren, die op hun beurt worden onderscheiden in één of meer concrete componenten. De beoordeling op deze (in totaal 20) componenten vindt plaats op een vierpunts-schaal: slecht – matig – redelijk – goed.

Inventarisatie van Kenmerken	
Evaluatieobject	
titel	
type object	
land(en), regio('s)	
financieel belang	
evaluatieperiode	
Evaluatieonderzoek	
gebruiksdoelstelling	
type onderzoek	
onderzoekuitgaven	€ (… %)
financieringsbron	
doorlooptijd	
Actoren	
opdrachtgever	
sturing of begeleiding	
evaluatoren	• •
kwalificaties	
selectiecriteria	

[36] Deze omvatten zowel PE's van individuele MFS-organisaties als gezamenlijke PE's ondernomen in het kader van een tijdelijk of permanent samenwerkingsverband.

Evaluatierapport	
datum	
auteur(s)	• •
onderzoeksvraagstelling	
ToR	
omvang	

Toelichting op invulling Inventarisatie van kenmerken

titel	de naam van het onderzoek volgens de definitieve versie van het eindrapport
type object	(bijv.) programma / project / sector / thema / instrument / strategie / organisatie
land(en), regio('s)	gebied(en) waar het onderzoek betrekking op heeft
financieel belang	het bedrag aan MFS-fondsen dat tijdens de onderzochte periode aan het evaluatieobject is besteed
evaluatieperiode	de periode die het onderzoek bestrijkt
gebruiksdoelstelling	(i) verantwoording en/of (ii) beleidsontwikkeling (leren) en/of (iii) managementbeslissing
type onderzoek	(bijv.) ex post / interim / evaluatie / review[37]/ evaluatie & formulering[38]
onderzoekuitgaven	in € (en in % van de financiële belang van het evaluatieobject)
financieringsbron	het budget(onderdeel) waaruit het onderzoek wordt gefinancierd
doorlooptijd	de periode die met de uitvoering van het onderzoek is gemoeid
opdrachtgever	functionaris of instantie die de ToR vaststelt en beslist over de aanvaarding van de eindrapportage
sturing/begeleiding	samenstelling van (eventueel ingesteld) gezelschap dat begeleiding of sturing gaf aan de evaluatie
evaluatoren	namen (+ eventuele werkkringen) van zowel teamleider als teamleden
kwalificaties	vermelding van de voornaamste opleidings- en ervaringsgegevens van de onderzoekers: ja / nee
selectiecriteria	zijn de criteria aan de hand waarvan de onderzoekers zijn geselecteerd expliciet vermeld (bijvoorbeeld in de ToR): ja / nee
datum	van de definitieve versie van het eindrapport
auteur(s)	namen (+ werkkringen) indien anderen dan de evaluatoren zelf
onderzoeksvraagstelling	vermelding c.q. herhaling (uit de ToR) van de onderzoekvragen in de hoofdtekst van het eindrapport: ja (p. ...) / nee
ToR	zijn deze – volledig of verkort – als bijlage in het rapport opgenomen? ja (volledig/verkort). p. ... / nee
omvang	van het rapport in pagina's (inclusief bijlagen)

37 review: 'lichte' vorm van evaluatieonderzoek, voornamelijk gebaseerd op reeds bestaand materiaal.

38 evaluatie & formulering: van deze combinatie is sprake als de onderzoeksopdracht zodanig expliciete aanwijzingen voor de opstelling van aanbevelingen ten behoeve van de voortzetting van de te evalueren activiteiten bevat, dat het risico bestaat dat de aanbevelingen de evaluatieve bevindingen gaan sturen in plaats van omgekeerd.

Beoordelingscriteria, indicatoren en componenten		
1	**Validiteit (meet het onderzoek daadwerkelijk wat het beoogt te meten?)**	**score**
1.1	**Probleemstelling**	
1.1.1	Helderheid van probleemstelling en uitwerking in onderzoekvragen	
1.1.2	Definitie van evaluatiecriteria	
1.2	**Evaluatieobject**	
1.2.1	Definitie, werking en afbakening van het onderzoeksobject	
1.2.2	Situering van het evaluatieobject in beleidsmatige en institutionele context	
1.3	**Beleidstheorie**	
1.3.1	Weergave of reconstructie van interventielogica en resultaatniveaus	
1.3.2	Operationalisering van resultaatmeting via indicatoren	
1.4	**Analyse**	
1.4.1	Gegevensbronnen, -verzameling en verwerking	
1.4.2	Onderbouwing van conclusies door bevindingen	
2	**Betrouwbaarheid (hoe stabiel zijn de onderzoekuitkomsten?)**	**score**
2.1	**Onderzoekmethoden**	
2.1.1	Specificatie en verantwoording van gehanteerde onderzoekmethoden	
2.1.2	Verificatie van gegevens / Triangulatie	
2.2	**Reikwijdte**	
2.2.1	Representativiteit van de steekproeftrekking c.q. casestudy-selectie	
2.2.2	Vermelding van beperkingen van het onderzoek	
2.3	**Onafhankelijkheid**	
2.3.1	Van het bronmateriaal (t.o.v. de belanghebbenden)	
2.3.2	Van de evaluatoren (t.o.v. de belanghebbenden)	
2.4	**Onderzoekverloop en kwaliteitsbewaking**	
2.4.1	Verantwoording van het onderzoekverloop	
2.4.2	Kwaliteitsbewaking via intern of extern toezicht	
3	**Bruikbaarheid (hoe goed zijn de onderzoekuitkomsten toepasbaar?)**	**score**
3.1	**Presentatie**	
3.1.1	Helderheid van de onderzoekdoelstelling	
3.1.2	Toegankelijkheid van de onderzoekresultaten	
3.2	**Aansluiting**	
3.2.1	Beantwoording van de onderzoeksvragen door de conclusies	
3.2.2	Uitvoerbaarheid van lessen of aanbevelingen	

Toelichting op invulling Beoordelingscriteria, -indicatoren en componenten

	validiteit
1.1.1	De probleemstelling formuleert kernachtig *waarop* (aan de hand van welke, met name genoemde, criteria) het evaluatieobject wordt beoordeeld. De onderzoeksvragen vormen de operationalisering van de probleemstelling.
1.1.2	Eenduidige omschrijving van de maatstaven – zoals doeltreffendheid – die worden aangelegd om het evaluatieobject te beoordelen.
1.2.1	Opsomming. omschrijving en begrenzing van de verzameling (operationele populatie) van onderzoekseenheden (naar type, doelgroep, locatie, periode, instelling, financiële omvang, enz.) waarop de onderzoeksresultaten betrekking hebben.
1.2.2	Weergave van relevante beleidsmatige achtergronden en uitgangspunten, evenals van het institutionele krachtenveld waarin het evaluatieobject opereert.
1.3.1	Uiteenzetting van de beleidstheorie, met de veronderstellingen over causale en finale relaties die aan onderzochte interventies ten grondslag heeft gelegen, en over de gehanteerde doel–middelenhiërarchie met de onderscheiden resultaatniveaus.
1.3.2	Mate waarin de indicatoren die op de verschillende resultaatniveaus zijn gedefinieerd, als specifiek, meetbaar en tijdgebonden kunnen worden beschouwd.
1.4.1	Zorgvuldigheid waarmee de gebruikte gegevensbronnen zijn geselecteerd, evenals de nauwkeurigheid en transparantie waarmee gegevens uit die bronnen worden geanalyseerd en verwerkt.
1.4.2	Mate waarin de conclusies daadwerkelijk worden gedekt door de onderzoekbevindingen.
	betrouwbaarheid
2.1.1	Nauwkeurige identificatie en rechtvaardiging van de gehanteerde onderzoekmethoden en technieken.
2.1.2	Mate waarin gegevens zijn gecontroleerd, en verschillende bronnen/methoden zijn gebruikt om informatie over dezelfde kenmerken en verschijnselen te verzamelen.
2.2.1	Mate waarin de conclusies uit de onderzochte steekproef c.q. van de uitgevoerde case studies gelden voor de hele onderzoekspopulatie.
2.2.2	Vermelding van en uitleg over (eventuele) tekortkomingen van het onderzoek en restricties aan de generaliseerbaarheid van de bevindingen en conclusies.
2.3.1	Mate waarin selectie en inhoud van geraadpleegde gegevensbronnen, in het bijzonder documentatie en respondenten, onafhankelijk waren van belanghebbenden bij de evaluatie zoals opdrachtgevers, uitvoerders en beneficianten.
2.3.2	Mate waarin de evaluatoren onafhankelijk opereerden en rapporteerden van belanghebbenden bij het onderzoek zoals opdrachtgevers, uitvoerders en beneficianten.
2.4.1	Beschrijving en verklaring van het verloop van de evaluatie, inclusief eventuele aanpassingen die ten opzichte van de oorspronkelijke opzet zijn aangebracht.
2.4.2	Controle op het ontwerp en/of de uitvoering van het onderzoek door een begeleidings- of stuurgroep binnen of buiten de MFS-organisatie(s).
	bruikbaarheid
3.1.1	Helderheid van de specificatie van het (buiten het onderzoek zelf gelegen – externe) doel van de evaluatie, *waarvoor* de onderzoekuitkomsten zullen worden of zijn gebruikt.

3.1.2	Duidelijkheid en volledigheid waarmee in het evaluatierapport en de samenvatting ervan de essentie van het onderzoek, en met name de hoofdbevindingen, zijn weergeven.
3.2.1	Volledigheid waarmee alle onderzoekvragen door de conclusies worden beantwoord.
3.2.2	Praktische uitvoerbaarheid van gepresenteerde aanbevelingen en de mate waarin deze binnen het bereik liggen van betrokken verantwoordelijken, vooral van de evaluatie-opdrachtgevers.

Bijlage 4 Conversie van MFP-b naar MFS-I beoordelingsscores

Gedurende de subsidieperiode van het Medefinancieringsprogramma-breed (MFP-b) (2003 – 2006) werd de methodologische kwaliteit van de door de MFO's uitgevoerde programma-evaluaties beoordeeld aan de hand van de tweede versie van een lijst[39] die bestond uit drie criteria: Validiteit, Betrouwbaarheid en Bruikbaarheid. Deze drie criteria waren samen geoperationaliseerd in 14 kwaliteitsindicatoren, waarbinnen in totaal 32 verschillende componenten werden onderscheiden.[40]

In het kader van de meta-evaluatie aan het einde van de MFP-breed periode bleek dat de hanteerbaarheid van de Beoordelingslijst voor verbetering vatbaar was. Dit werd in eerste instantie bereikt door componenten die logisch – en praktisch – sterk samenhangen, te combineren. Vervolgens zijn beoordelingscomponenten geschrapt waarvan in de praktijk was gebleken dat hun onderscheidend vermogen te gering was om de "kosten" van hun toepassing te rechtvaardigen. Op deze manier kon de Beoordelingslijst zowel worden gestroomlijnd als aanzienlijk ingekort.

Alle aanpassingen werden bij de aanvang van de nieuwe (MFS-I) subsidieperiode (en uiteraard voordat de eerste studies ter beoordeling werden aangeboden) doorgevoerd. De nieuwe (derde) versie van de lijst[41] die zo ontstond, bevat nog steeds de drie hoofdcriteria Validiteit, Betrouwbaarheid en Bruikbaarheid, maar deze zijn nu geoperationaliseerd in 10 indicatoren, die op hun beurt verder zijn geconcretiseerd in 20 componenten. De tweede en derde versies van de lijst zijn toegevoegd als bijlagen 4a en 4b.

Om de kwaliteitsbeoordelingen uit de twee opeenvolgende periodes vergelijkbaar te houden moesten de scores van de tweede versie van de Beoordelingslijst worden geconverteerd naar die van de derde versie. De onderstaande tabel laat zien dat zeven componenten zijn geschrapt, terwijl negen componenten uit de bestaande lijst werden gecombineerd tot vier nieuwe. Ook leek het bij nader inzien logischer sommige componenten te rangschikken onder andere hoofdcriteria. Dat is in drie gevallen gebeurd.

De vierpuntsschaal dwingt tot scherpe keuzes in het belang van de transparantie van het kwaliteitsoordeel op de verschillende onderdelen. Tegenover het voordeel van de overzichtelijkheid staat evenwel het nadeel van een zeker verlies aan nuanceringsmogelijkheden in die gevallen waarin men daarover zou wensen te beschikken. De gehanteerde schaal laat niet toe in geval van twijfel tussen twee scores een tussenliggende waardering te kiezen. In dergelijke gevallen is aan de betreffende programmaevaluatie het voordeel van de twijfel gegeven door de *hogere* score te kiezen. Hierdoor kan dus een zekere positieve vertekening

[39] *Beoordelingslijst Programma-evaluaties*, versie maart 2005. Deze lijst was een aanpassing van de eerste – experimentele – versie: *Analyse van MFP Programma-evaluaties*, die in april 2004 was ontwikkeld om tijdens de aanloopperiode ervaring op te doen met de analyse en beoordeling van de eerste aangeleverde studies.

[40] Zie *de Kwaliteit van Programma-evaluaties in het Medefinancieringsprogramma 2003–2006*, IOB juni 2006, pp. 11-12.

[41] *Beoordelingslijst Programma-evaluaties*, versie januari 2007.

in het beoordelingsbeeld ontstaan. Die is niet op een andere manier te voorkomen dan door haar in te ruilen voor het risico van een negatieve vertekening, een “oplossing” die meer nadelen vertoont dan het probleem zelf.

Tegen deze achtergrond is tijdens de conversie van de oude naar de nieuwe beoordelingslijst bij het combineren van scores die onderling verschilden, voor de *lagere* waardering gekozen. Op die manier is vermeden dat een aan de beoordelingssystematiek reeds inherent vertekeningsrisico door het conversieproces verder wordt versterkt.
In de onderstaande tabellen is de conversie van de componenten van de twee Beoordelingslijsten in beide richtingen weergegeven. Daarom bevat de eerste tabel in de linker kolom 32 scoreregels en de tweede slechts 20.

Conversie Scores PE Beoordelingslijsten van MFP-b naar MFS-I Model			
Model MFP-b		**Model MFS-I**	
1.1.1	vermelding aanleiding evaluatie	---	- vervallen -
1.1.2	duidelijkheid gebruiksdoelstelling	3.1.1	helderheid onderzoekdoelstelling
1.2.1	helderheid probleemstelling	1.1.1	helderheid probleemstelling – uitwerking onderzoekvragen
1.2.2	uitwerking onderzoekvragen		
1.2.3	definiëring evaluatiecriteria	1.1.2	definitie evaluatiecriteria
1.3.1	afbakening onderzoekobject	1.2.1	werking en afbakening onderzoeksobject
1.3.2	opzet en wijze van uitvoering		
1.3.3	situering object context	1.2.2	situering object in context
1.4.1	reconstructie interventielogica	1.3.1	weergave / reconstructie interventie-logica en resultaatniveaus
1.4.2	onderscheiding resultaatniveaus		
1.4.3	operationalisering via indicatoren	1.3.2	operationalisering via indicatoren
1.5.1	specificatie onderzoekmethoden	2.1.1	specificatie en verantwoording onderzoekmethoden
1.5.2	verantwoording onderzoekmethoden		
1.5.3	vermelding onderzoekbeperkingen	2.2.2	vermelding onderzoekbeperkingen
1.6.1	gegevensanalyse	1.4.1	gegevensverzameling en -verwerking
1.6.2	onderbouwing conclusies	1.4.2	onderbouwing conclusies
1.6.3	aansluiting aanbevelingen-conclusies	---	- vervallen -
2.1.1	identificatie gegevensbronnen	2.1.1	specificatie en verantwoording onderzoekmethoden
2.1.2	triangulatie	2.1.2	triangulatie
2.2.1	steekproef / casestudy-selectie	2.2.1	representativiteit
2.3.1	onafhankelijkheid gegevensbronnen	2.3.1	onafhankelijkheid gegevensbronnen
2.3.2	onafhankelijkheid onderzoekers	2.3.2	onafhankelijkheid evaluatoren
2.4.1	uitvoering binnen planning / budgettering	---	- vervallen -

2.4.2	verantwoording onderzoekverloop	2.4.1	verantwoording onderzoekverloop
2.5.1	verwerking commentaar	---	- vervallen -
2.5.2	begeleidings- of stuurgroep	2.4.2	kwaliteitstoezicht
3.1.1	toegankelijkheid onderzoekresultaten	3.1.2	toegankelijkheid onderzoekresultaten
3.1.2	realisme aanbevelingen	3.2.2	uitvoerbaarheid aanbevelingen
3.2.1	beantwoording onderzoekvragen	3.2.1	beantwoording onderzoekvragen
3.2.2	verschillenverklaring	---	- vervallen -
3.3.1	helderheid samenvatting	---	- vervallen -
3.3.2	onderscheiding aanbeveling / conclusies	---	- vervallen -

Conversie Scores PE Beoordelingslijsten van MFS-I naar MFP-b Model			
Model MFS-I		**Model MFP-b**	
1.1.1	helderheid probleemstelling – uitwerking onderzoekvragen	1.2.1	helderheid probleemstelling
		1.2.2	uitwerking onderzoekvragen
1.1.2	definitie evaluatiecriteria	1.2.3	definiëring evaluatiecriteria
1.2.1	werking en afbakening onderzoekobject	1.3.1	afbakening onderzoekobject
		1.3.2	opzet en wijze van uitvoering
1.2.2	situering object in context	1.3.3	situering object in context
1.3.1	weergave / reconstructie interventielogica en resultaatniveaus	1.4.1	reconstructie interventielogica
		1.4.2	onderscheiding resultaatniveaus
1.3.2	operationalisering via indicatoren	1.4.3	operationalisering via indicatoren
1.4.1	gegevensverzameling en -verwerking	1.6.1	gegevensanalyse
1.4.2	onderbouwing conclusies	1.6.2	onderbouwing conclusies
2.1.1	specificatie en verantwoording onderzoekmethoden	1.5.1	specificatie onderzoekmethoden
		1.5.2	verantwoording onderzoekmethoden
		2.1.1	identificatie gegevensbronnen
2.1.2	triangulatie	2.1.2	triangulatie
2.2.1	representativiteit	2.2.1	steekproef / casestudy-selectie
2.2.2	vermelding onderzoekbeperkingen	1.5.3	vermelding onderzoekbeperkingen
2.3.1	onafhankelijkheid gegevensbronnen	2.3.1	onafhankelijkheid gegevensbronnen
2.3.2	onafhankelijkheid evaluatoren	2.3.2	onafhankelijkheid onderzoekers
2.4.1	verantwoording onderzoekverloop	2.4.2	verantwoording onderzoekverloop
2.4.2	kwaliteitstoezicht	2.5.2	begeleidings- of stuurgroep
3.1.1	helderheid onderzoekdoelstelling	1.1.2	duidelijkheid gebruiksdoelstelling
3.1.2	toegankelijkheid onderzoekresultaten	3.1.1	toegankelijkheid onderzoekresultaten
3.2.1	beantwoording onderzoekvragen	3.2.1	beantwoording onderzoekvragen
3.2.2	uitvoerbaarheid aanbevelingen	3.1.2	realisme aanbevelingen

Bijlage 4a Beoordelingslijst Programma-evaluaties

MFP-breed Model (versie maart 2005)

Deze beoordelingslijst is gehanteerd om de opzet en uitvoering van Programma-evaluaties (PE's)[42] binnen het Medefinancieringsprogramma (MFP-breed) te toetsen aan de kwaliteitseisen van de Regeling Prestatiegegevens Evaluatieonderzoek (RPE) van de Rijksoverheid.

Inventarisatie van Kenmerken	
Evaluatieobject	
titel	
type object	
land(en), regio('s)	
financiële omvang	
evaluatieperiode	
Evaluatieonderzoek	
gebruiksdoelstelling	
soort onderzoek	
onderzoekuitgaven in € (en in %)	€ (… %)
financieringsbron	
doorlooptijd	
Actoren	
opdrachtgever	
sturing of begeleiding	
evaluatoren	• •
vermelding kwalificaties	
explicitering selectiecriteria	

[42] Deze omvatten zowel individuele PE's van afzonderlijke medefinancieringsorganisaties (MFO's) als gezamenlijke PE's ondernomen in het kader van de samenwerking in het Medefinancieringsprogramma-Breed Netwerk (MBN).

Evaluatierapport	
datum	
auteur(s)	
onderzoeksvraagstelling	
ToR	
omvang	

Beoordelingscriteria, indicatoren en scores		
1	**Validiteit (meet men wat men beoogt te meten?)**	**score**
1.1	**Aanleiding en doelstelling**	
1.1.1	Vermelding van de aanleiding voor de evaluatie	
1.1.2	Duidelijkheid van de (onderzoek)gebruiksdoelstelling	
1.2	**Probleemstelling en onderzoekvragen**	
1.2.1	Helderheid van de probleemstelling	
1.2.2	Uitwerking van de probleemstelling in relevante onderzoekvragen	
1.2.3	Definiëring van de gehanteerde evaluatiecriteria	
1.3	**Evaluatieobject en context**	
1.3.1	Definiëring en afbakening van het onderzoekobject	
1.3.2	Beschrijving van de organisatorische opzet en wijze van uitvoering	
1.3.3	Situering van het evaluatieobject in beleidsmatige en institutionele context	
1.4	**Interventielogica**	
1.4.1	Weergave dan wel reconstructie van de interventielogica	
1.4.2	Onderscheiding van verschillende resultaatniveaus	
1.4.3	Operationalisering van resultaatmeting door middel van indicatoren	
1.5	**Onderzoeksmethoden**	
1.5.1	Specificatie van de gehanteerde onderzoekmethoden	
1.5.2	Verantwoording van de toegepaste methoden	
1.5.3	Vermelding van de beperkingen van het onderzoek	
1.6	**Consistentie en solliditeit**	
1.6.1	Gegevensanalyse en -interpretatie, formulering van bevindingen	
1.6.2	Onderbouwing van conclusies door bevindingen	
1.6.3	Overeenstemming tussen aanbevelingen en conclusies	

2	Betrouwbaarheid (zijn de onderzoekuitkomsten stabiel?)	score
2.1	**Gegevensbronnen**	
2.1.1	Identificatie van geraadpleegde gegevensbronnen	
2.1.2	Verificatie van hun betrouwbaarheid en nauwkeurigheid / Triangulatie	
2.2	**Representativiteit**	
2.2.1	Verantwoording van de onderzochte steekproef c.q. casestudy selectie	
2.3	**Onafhankelijkheid**	
2.3.1	Onafhankelijkheid van de gegevensbronnen t.o.v. de belanghebbenden	
2.3.2	Onafhankelijkheid van de onderzoekers t.o.v. de belanghebbenden	
2.4	**Onderzoekverloop**	
2.4.1	Uitvoering van de evaluatie binnen de geplande tijd en het beschikbare budget	
2.4.2	Verantwoording van het onderzoekverloop	
2.5	**Kwaliteitsbewaking**	
2.5.1	Verwerking van commentaar van belanghebbenden	
2.5.2	“Intern” dan wel “extern” toezicht door begeleidings- of stuurgroep	
3	**Bruikbaarheid (zijn de onderzoekuitkomsten toepasbaar?)**	**score**
3.1	**Aansluiting**	
3.1.1	Eenduidigheid en toegankelijkheid van de onderzoekresultaten	
3.1.2	Operationaliteit en realisme van aanbevelingen of lessen	
3.2	**Volledigheid**	
3.2.1	Beantwoording van de onderzoekvragen door de conclusies	
3.2.2	Verklaring van verschillen tussen beoogde en behaalde resultaten	
3.3	**Presentatie**	
3.3.1	Helderheid en representativiteit van de samenvatting	
3.3.2	(Onder)scheiding van aanbevelingen en conclusies	

Bijlage 4b Beoordelingslijst Programma-evaluaties

MFS-I Model (versie januari 2007)

Deze beoordelingslijst wordt gehanteerd om de opzet en uitvoering van Programma-evaluaties (PE's)[43] binnen het Medefinancieringsstelsel (MFS) te toetsen aan gangbare kwaliteitseisen voor sociaalwetenschappelijk onderzoek.
Daartoe wordt eerst een aantal basiskenmerken van de betreffende PE geregistreerd. Deze kenmerken hebben achtereenvolgens betrekking op het geëvalueerde object, het uitgevoerde onderzoek, de daarbij betrokken actoren en het onderzoeksrapport. Vervolgens vindt de toetsing plaats aan de drie kwaliteitscriteria: Validiteit, Betrouwbaarheid en Bruikbaarheid. Deze criteria zijn geoperationaliseerd door de definiëring van een aantal indicatoren, die op hun beurt worden onderscheiden in één of meer concrete componenten. De beoordeling op deze (in totaal 20) componenten vindt plaats op een vierpunts-schaal: slecht – matig – redelijk – goed.

Inventarisatie van Kenmerken	
Evaluatieobject	
titel	
type object	
land(en), regio('s)	
financieel belang	
evaluatieperiode	
Evaluatieonderzoek	
gebruiksdoelstelling	
type onderzoek	
onderzoekuitgaven	€ (… %)
financieringsbron	
doorlooptijd	
Actoren	
opdrachtgever	
sturing of begeleiding	
evaluatoren	• •
kwalificaties	
selectiecriteria	

[43] Deze omvatten zowel afzonderlijke PE's van individuele MFS-organisaties als gezamenlijke PE's ondernomen in het kader van een tijdelijk of permanent samenwerkingsverband.

Evaluatierapport	
datum	
auteur(s)	• •
onderzoeksvraagstelling	
ToR	
omvang	

Beoordelingscriteria, indicatoren en componenten		
1	**Validiteit (meet het onderzoek daadwerkelijk wat het beoogt te meten?)**	**score**
1.1	**Probleemstelling**	
1.1.1	Helderheid van probleemstelling en uitwerking in onderzoekvragen	
1.1.2	Definitie van evaluatiecriteria	
1.2	**Evaluatieobject**	
1.2.1	Definitie, werking en afbakening van het onderzoekobject	
1.2.2	Situering van het evaluatieobject in beleidsmatige en institutionele context	
1.3	**Beleidstheorie**	
1.3.1	Weergave of reconstructie van interventielogica en resultaatniveaus	
1.3.2	Operationalisering van resultaatmeting via indicatoren	
1.4	**Analyse**	
1.4.1	Gegevensbronnen, -verzameling en verwerking	
1.4.2	Onderbouwing van conclusies door bevindingen	
2	**Betrouwbaarheid (hoe stabiel zijn de onderzoekuitkomsten?)**	**score**
2.1	**Onderzoekmethoden**	
2.1.1	Specificatie en verantwoording van gehanteerde onderzoeksmethoden	
2.1.2	Verificatie van gegevens / Triangulatie	
2.2	**Reikwijdte**	
2.2.1	Representativiteit van de steekproeftrekking c.q. casestudy-selectie	
2.2.2	Vermelding van beperkingen van het onderzoek	
2.3	**Onafhankelijkheid**	
2.3.1	Van het bronmateriaal (t.o.v. de belanghebbenden)	
2.3.2	Van de evaluatoren (t.o.v. de belanghebbenden)	
2.4	**Onderzoekverloop en kwaliteitsbewaking**	
2.4.1	Verantwoording van het onderzoekverloop	
2.4.2	Kwaliteitsbewaking via intern of extern toezicht	
3	**Bruikbaarheid (hoe goed zijn de onderzoeksuitkomsten toepasbaar?)**	**score**
3.1	**Presentatie**	
3.1.1	Helderheid van de onderzoekdoelstelling	
3.1.2	Toegankelijkheid van de onderzoekresultaten	
3.2	**Aansluiting**	
3.2.1	Beantwoording van de onderzoekvragen door de conclusies	
3.2.2	Uitvoerbaarheid van lessen of aanbevelingen	

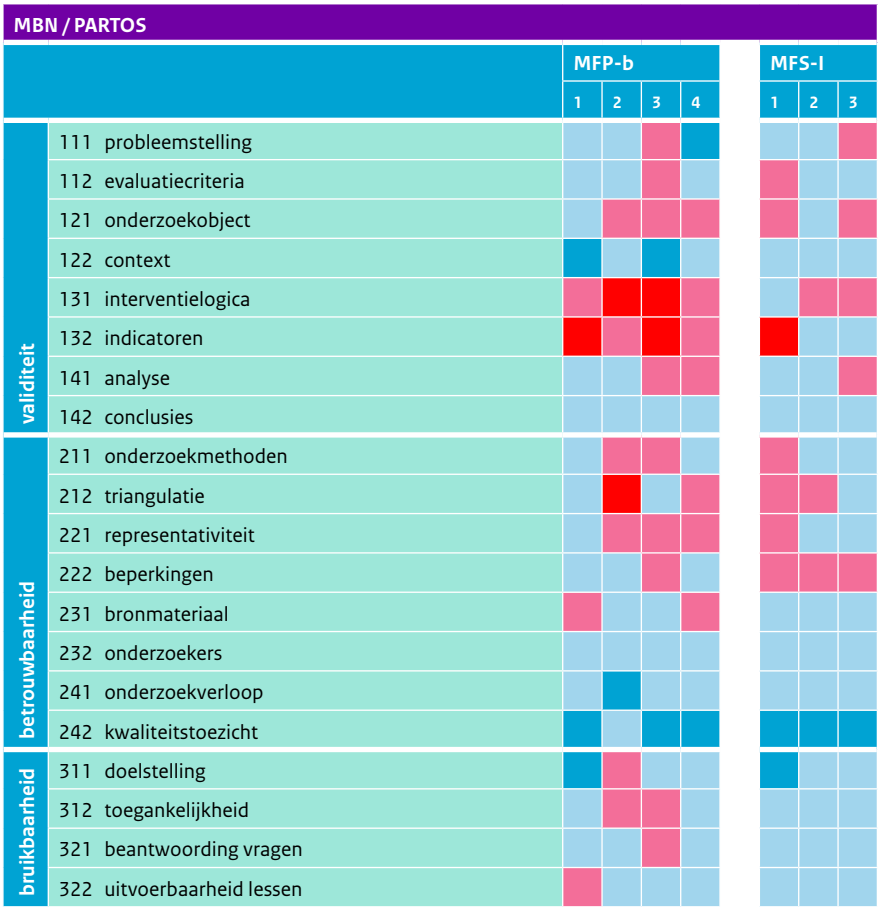
Bijlage 5 Scorekaarten

Deze bijlage bevat twaalf scorekaarten: één voor elk van de MFS-organisaties waarvan de programma-evaluaties op hun kwaliteiten zijn beoordeeld en een twaalfde voor de gezamenlijk ondernomen PE's in MBN en later Partos-verband.

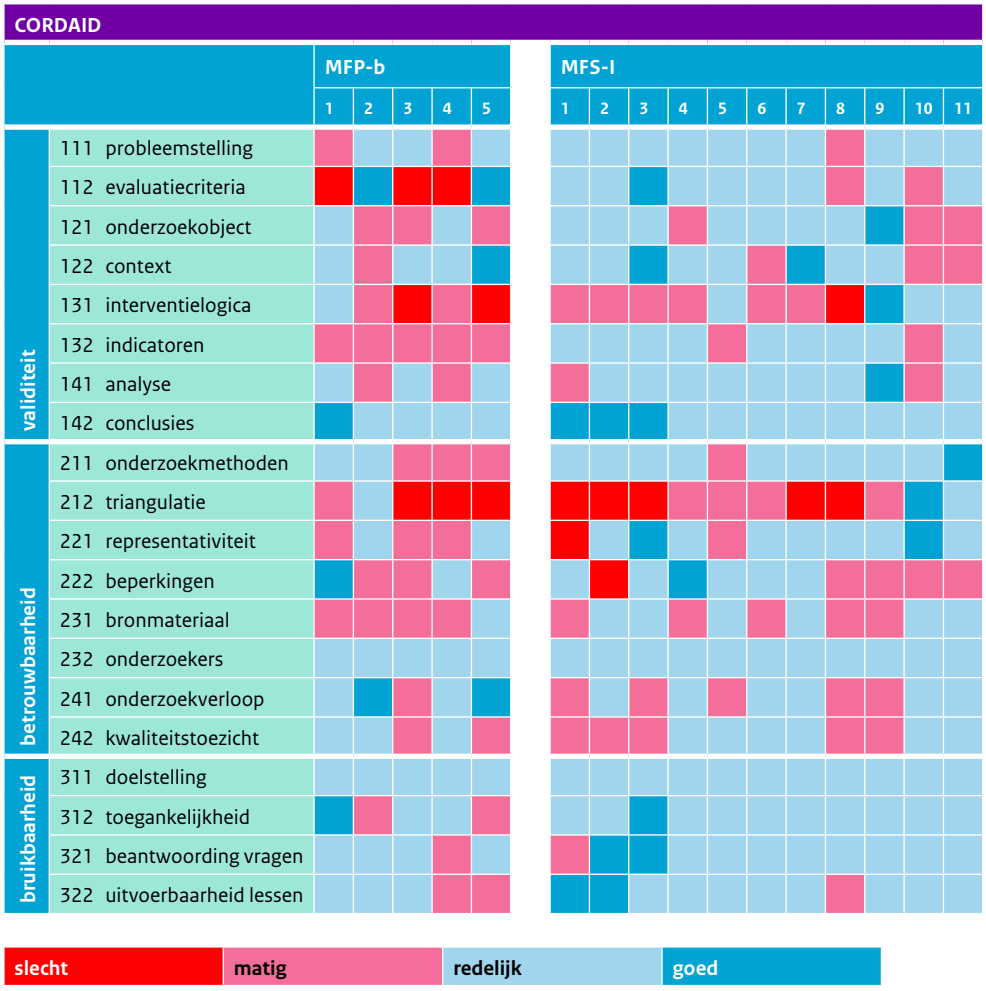
Elke kaart bevat de scores van alle programma-evaluaties die gedurende de MFP-breed (26) en MFS-I (72) subsidieperiodes op elk van de 20 beoordelingscomponenten. De volgorde van de beoordeelde PE's is die waarin hun eindrapporten zijn gedateerd. Onder de scoretabellen zijn in aparte lijsten de titels van de genummerde PE's weergegeven.

Linksonder wordt de dekkingsgraad getoond die de organisatie met de op de scorekaart vermelde PE's gedurende het MFS-I heeft behaald. Voor de berekening van de dekkingsgraad wordt verwezen naar § 5.1. De vermelde bedragen werden overgenomen uit Tabel 5–l.

Rechtsonder ten slotte is de score vermeld die aan de betreffende subsidieontvanger is toegekend op het criterium *Lerend vermogen in de keten* als onderdeel van de Organisatietoets ten behoeve van de toelating tot het MFS-I. De toelichting op de relevantie hiervan vindt men in § 5.4.1.



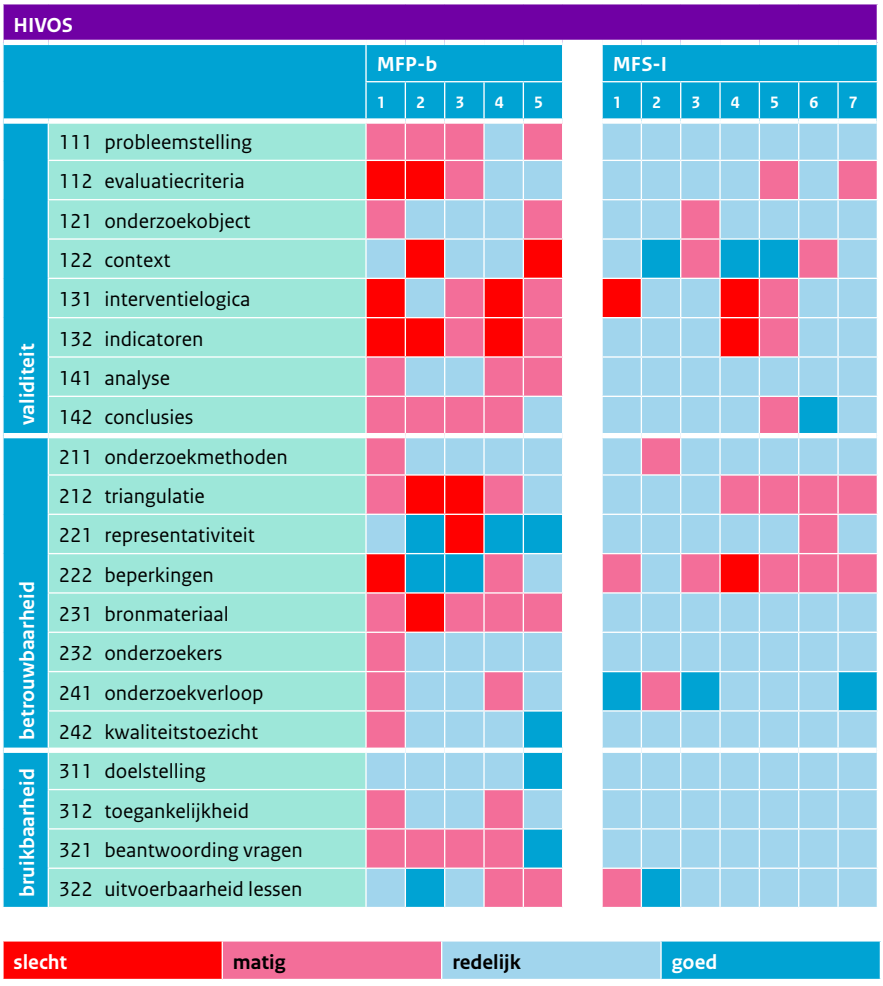
#	MFP-b: programmaevaluaties	#	MFS-I: programmaevaluaties
1	Vrouwenorganisaties in maatschappij-opbouw	1	Conflict Transformation: a Science and an Art
2	Gezondheidsthema	2	Microfinanciering
3	Assessing Civil Society Participation	3	Inheemse Volken
4	HIV/AIDS		


CORDAID
MFP-b
1 2 3 4 5
MFS-I
1 2 3 4 5 6 7 8 9 10 11
validiteit
111 probleemstelling
112 evaluatiecriteria
121 onderzoekobject
122 context
131 interventielogica
132 indicatoren
141 analyse
142 conclusies
betrouwbaarheid
211 onderzoekmethoden
212 triangulatie
221 representativiteit
222 beperkingen
231 bronmateriaal
232 onderzoekers
241 onderzoekverloop
242 kwaliteitstoezicht
bruikbaarheid
311 doelstelling
312 toegankelijkheid
321 beantwoording vragen
322 uitvoerbaarheid lessen
slecht
matig
redelijk
goed

#	MFP-b: programmaevaluaties	3	Biologische Fruitteelt Latijns-Amerika
1	Regional and Support Organisaton Programme	4	Kinderprogramma
2	Evaluative Enquiry of ICCO/Cordaid Consult	5	Midden-Oosten en Midden- en Oost-Europa Programma
3	ASA Bangladesh	6	Capaciteitsversterking Financieel Management
4	Sustainable Agriculture Programme of VECO Indonesia	7	Sloppenwijkbewoners
5	Programma Suriname	8	Performance-Based Financing
#	MFS-I: programmaevaluaties	9	SME Support and Value Chain Development
1	Common Programming and Shared Indicators	10	Disaster Risk Reduction
2	PHC Support Programme in Cyangugu Province	11	HIV Programma

programma-evaluatiedekking mfs-I	
mfs-I subsidiebedrag	€ 396.736.000
object programma-evaluaties	€ 151.249.348
dekking	**38 %**

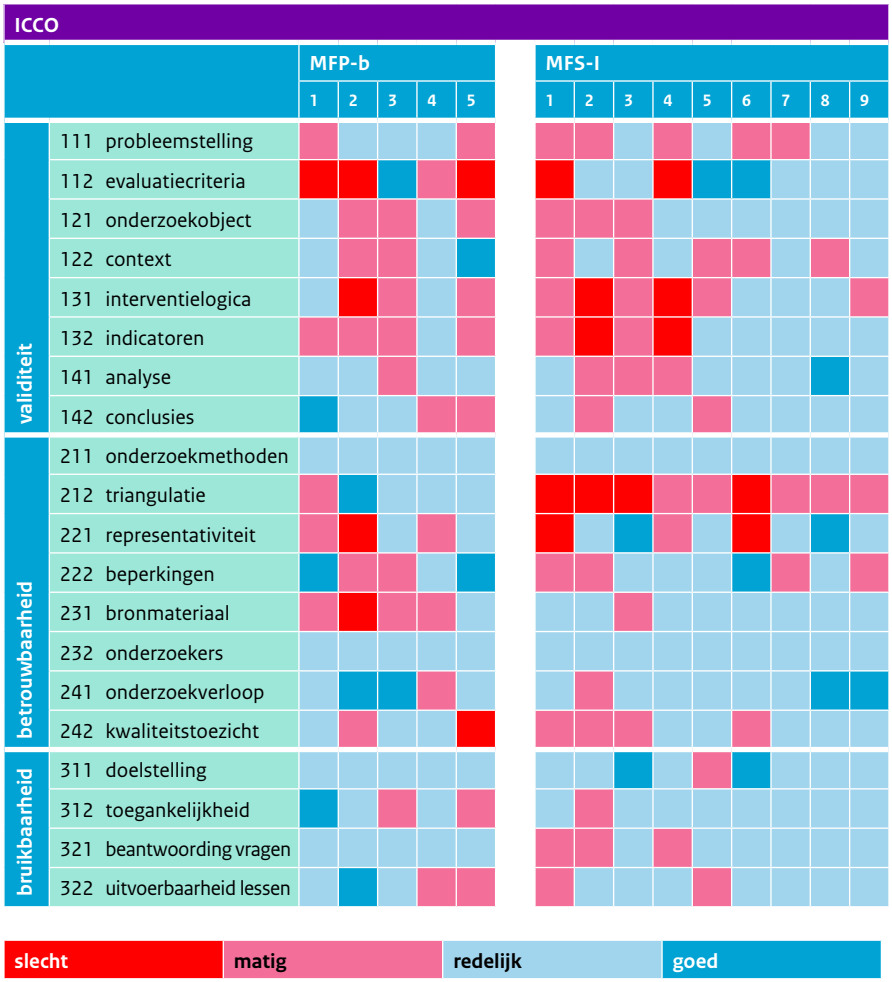
ex ante kwaliteitsbeoordeling evaluatiefunctie	
(organisatie-toets voor subsidietoekenning mfs-I, 2006)	
Cordaid-score *lerend vermogen in de keten*	9
percentage van maximum score	**100 %**


HIVOS
MFP-b
1 2 3 4 5
MFS-I
1 2 3 4 5 6 7
validiteit
111 probleemstelling
112 evaluatiecriteria
121 onderzoekobject
122 context
131 interventielogica
132 indicatoren
141 analyse
142 conclusies
betrouwbaarheid
211 onderzoekmethoden
212 triangulatie
221 representativiteit
222 beperkingen
231 bronmateriaal
232 onderzoekers
241 onderzoekverloop
242 kwaliteitstoezicht
bruikbaarheid
311 doelstelling
312 toegankelijkheid
321 beantwoording vragen
322 uitvoerbaarheid lessen
slecht
matig
redelijk
goed

#	MFP-b: programmaevaluaties	#	MFS-I: programmaevaluaties
1	Review of the Hivos Central Asia Programme	1	LGBT Emancipation in Southern Africa
2	Agrobiodiversity in Action	2	Missing Links: Organic Chains
3	Access for All: Equal Opportunities in Cyberspace	3	Evaluación del Programa Arte y Cultura
4	Film Sector in Asia and Southeastern Europe	4	Tsunami Response
5	Ondersteuning Milieubeweging Latijns-Amerika	5	Human Rights and Democratisation in Kenya
		6	ICT & Media Programme
		7	HIV & AIDS Programme

programma-evaluatiedekking mfs-I	
mfs-I subsidiebedrag	€ 231.995.000
object programma-evaluaties	€ 113.231.074
dekking	**49 %**

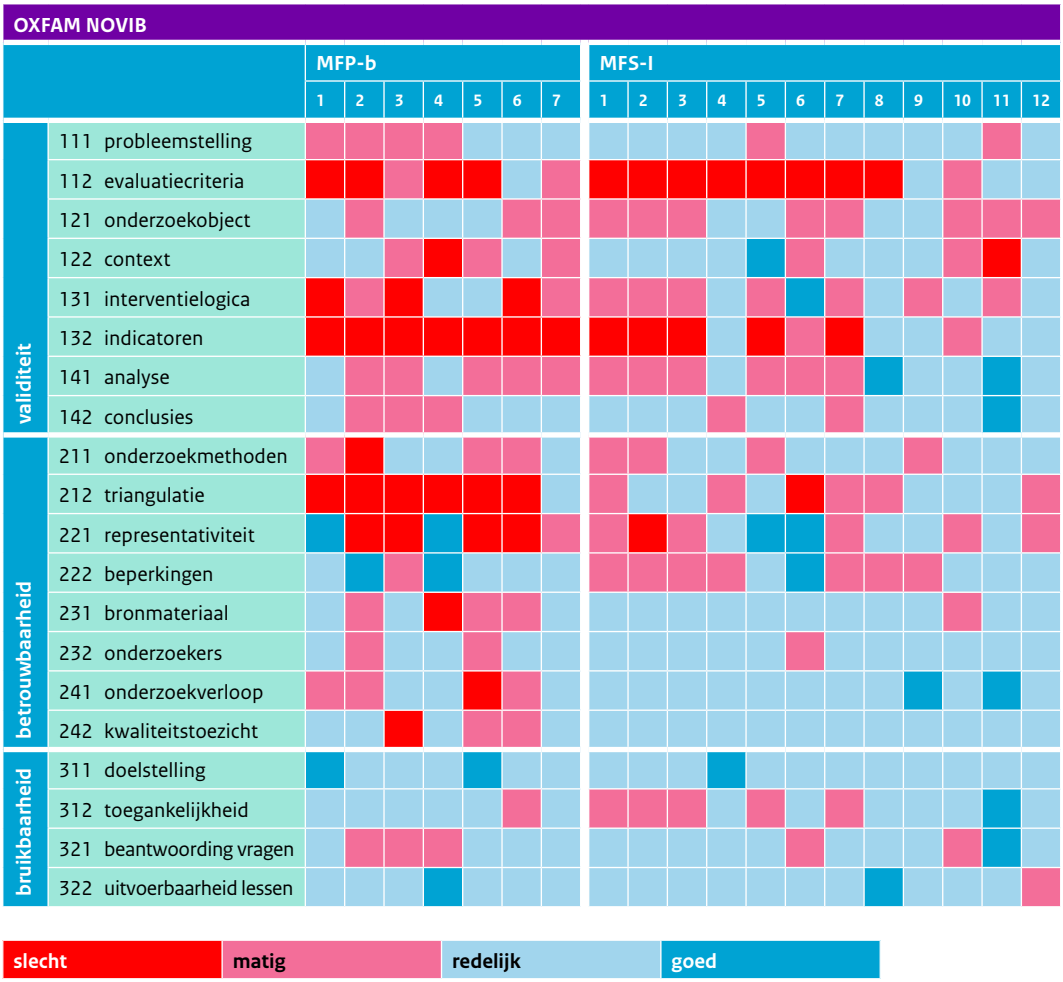
ex ante kwaliteitsbeoordeling evaluatiefunctie	
(organisatie-toets voor subsidietoekenning mfs-I, 2006)	
Hivos-score *lerend vermogen in de keten*	9
percentage van maximum score	**100 %**


ICCO
MFP-b
1 2 3 4 5
MFS-I
1 2 3 4 5 6 7 8 9
validiteit
111 probleemstelling
112 evaluatiecriteria
121 onderzoekobject
122 context
131 interventielogica
132 indicatoren
141 analyse
142 conclusies
betrouwbaarheid
211 onderzoekmethoden
212 triangulatie
221 representativiteit
222 beperkingen
231 bronmateriaal
232 onderzoekers
241 onderzoekverloop
242 kwaliteitstoezicht
bruikbaarheid
311 doelstelling
312 toegankelijkheid
321 beantwoording vragen
322 uitvoerbaarheid lessen
slecht
matig
redelijk
goed

#	MFP-b: programmaevaluaties	2	Kameroen Programma
1	Regional and Support Organisaton Programme	3	ACT Lobbycampagne Indonesië
2	Strategische Samenwerking ICCO-IKV 2002-2003	4	Strong People, Strong State
3	Evaluative Enquiry of ICCO/Cordaid Consult	5	Technical and Vocational Education and Training
4	Organisational Development supported by ICCO	6	Terrafina Microfinance
5	Country Evaluation Philippines	7	International Markets Programme
#	**MFS-I: programmaevaluaties**	8	Local Market Development Programme
1	Stille Oceaan Programma	9	Food Security Programme

programma-evaluatiedekking	
a: mfs-I subsidiebedrag	€ 443.961.737
b: totale object programma-evaluaties	€ 193.454.805
dekking: b/a	**44 %**

ex ante kwaliteitsbeoordeling evaluatiefunctie	
(organisatie-toets subsidietoekenning 2006)	
ICCO-score *lerend vermogen in de keten*	8
percentage van maximum score	**89 %**


OXFAM NOVIB
MFP-b
MFS-I
1 2 3 4 5 6 7
1 2 3 4 5 6 7 8 9 10 11 12
validiteit
111 probleemstelling
112 evaluatiecriteria
121 onderzoekobject
122 context
131 interventielogica
132 indicatoren
141 analyse
142 conclusies
betrouwbaarheid
211 onderzoekmethoden
212 triangulatie
221 representativiteit
222 beperkingen
231 bronmateriaal
232 onderzoekers
241 onderzoekverloop
242 kwaliteitstoezicht
bruikbaarheid
311 doelstelling
312 toegankelijkheid
321 beantwoording vragen
322 uitvoerbaarheid lessen
slecht
matig
redelijk
goed

#	MFP-b: programmaevaluaties	3	Core Country Programme Evaluation India
1	Violence against Women in South Africa	4	Review Linkis Oxfam Novib – Draagvlak-versterking in NL
2	Oxfam Mekong Initiative (OMI) Review	5	Core Country Programme Evaluation Angola
3	Joint Oxfam Programme Malawi (JOPM)	6	Humanitarian Capacity Building
4	Agrobiodiversity in Action	7	Core Country Programme Evaluation Palestinians in the ME
5	Oxfam International's Coffee Campaign	8	Moving a Mountain: Stop EPA Campaign
6	Joint Oxfam HIV and AIDS Programme in South Africa	9	Core Country Programme Evaluation Ethiopia
7	Food Security Policy and Practices	10	Global Programme
#	MFS-I: programmaevaluaties	11	Core Country Programme Evaluation Burundi
1	Core Country Programme Evaluation Nigeria	12	Evaluación del Programa de América Latina
2	Core Country Programme Evaluation Somalia		

programma-evaluatiedekking mfs-I	
mfs-I subsidiebedrag	€ 438.946.000
object programma-evaluaties	€ 238.752.538
dekking	**54 %**

ex ante kwaliteitsbeoordeling evaluatiefunctie	
(organisatie-toets voor subsidietoekenning mfs-I, 2006)	
Novib-score *lerend vermogen in de keten*	8
percentage van maximum score	**89 %**

ETC		MFS-I 1	MFS-I 2
validiteit	111 probleemstelling		
	112 evaluatiecriteria		
	121 onderzoekobject		
	122 context		
	131 interventielogica		
	132 indicatoren		
	141 analyse		
	142 conclusies		
betrouwbaarheid	211 onderzoekmethoden		
	212 triangulatie		
	221 representativiteit		
	222 beperkingen		
	231 bronmateriaal		
	232 onderzoekers		
	241 onderzoekverloop		
	242 kwaliteitstoezicht		
bruikbaarheid	311 doelstelling		
	312 toegankelijkheid		
	321 beantwoording vragen		
	322 uitvoerbaarheid lessen		

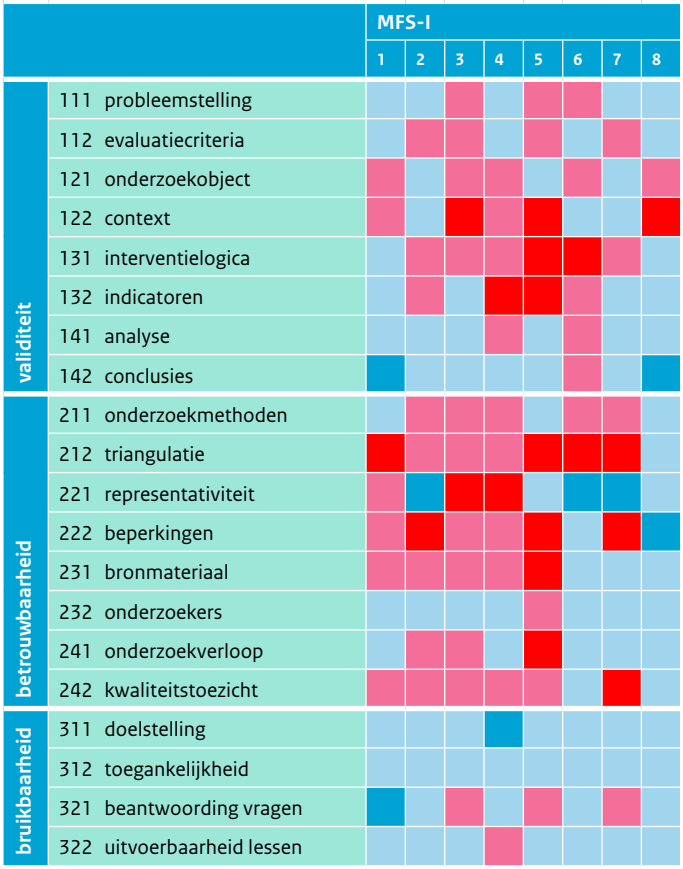
slecht	matig	redelijk	goed

#	MFS-I: programmaevaluatie	#	MFS-I: programmaevaluatie
1	PROLINNOVA	2	Energy Access Programme

programma-evaluatiedekking	
a: mfs-I subsidiebedrag	€ 12.987.552
b: totale object programma-evaluaties	€ 5.152.724
dekking: b/a	**40 %**

ex ante kwaliteitsbeoordeling evaluatiefunctie	
(organisatie-toets subsidietoekenning 2006)	
ETC-score *lerend vermogen in de keten*	8
percentage van maximum score	**89 %**

INTERNATIONAL CHILD SUPPORT

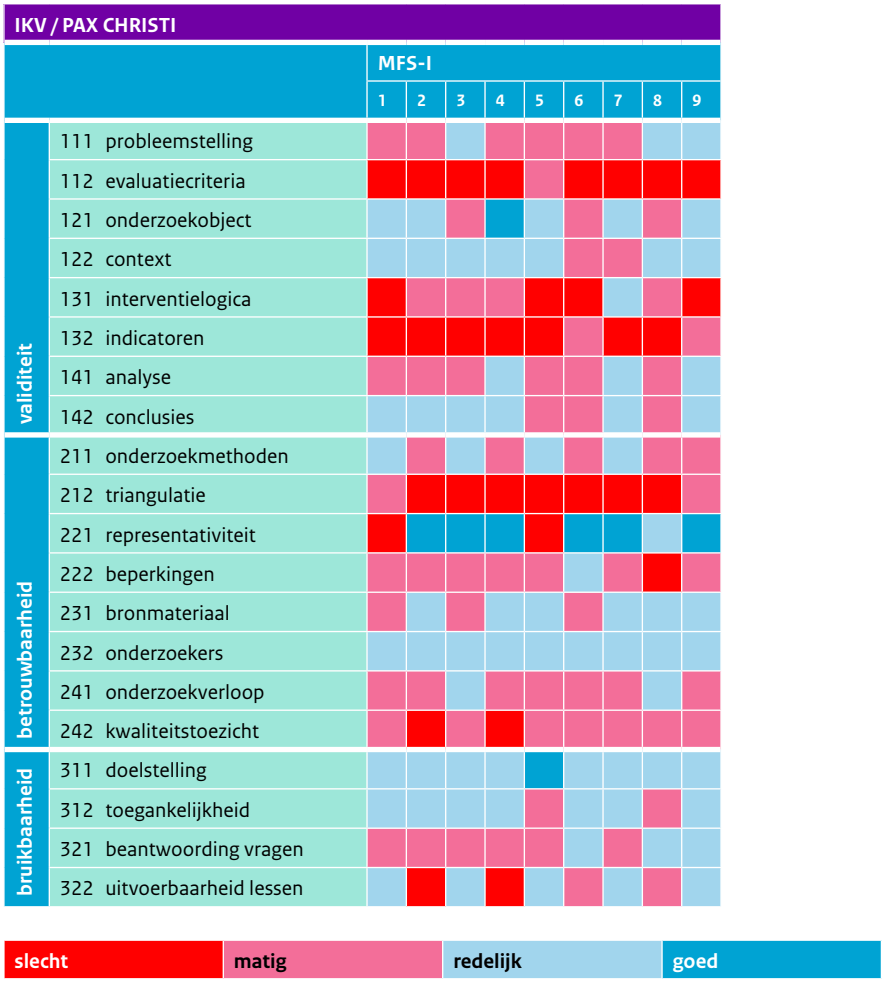


slecht	matig	redelijk	goed

#	MFS-I: programmaevaluaties	#	MFS-I: programmaevaluaties
1	Rice Bank Evaluation	5	African Child Policy Forum (ACPF)
2	Action for Children Programme	6	CDCD Programme Cambodia
3	Aflatoun Secretariat	7	Change for Children (CfC)
4	Knowledge Centre on Civic Driven Child Development (CDCD)	8	Child Rights in School Improvement Programme

programma-evaluatiedekking	
a: mfs-I subsidiebedrag	€ 21.433.634
b: totale object programma-evaluaties	€ 11.461.476
dekking: b/a	**53 %**

ex ante kwaliteitsbeoordeling evaluatie-functie	
(organisatie-toets subsidietoekenning 2006)	
ICS-score *lerend vermogen in de keten*	7
percentage van maximum score	**78 %**


IKV / PAX CHRISTI
MFS-I
1 2 3 4 5 6 7 8 9
validiteit
111 probleemstelling
112 evaluatiecriteria
121 onderzoekobject
122 context
131 interventielogica
132 indicatoren
141 analyse
142 conclusies
betrouwbaarheid
211 onderzoekmethoden
212 triangulatie
221 representativiteit
222 beperkingen
231 bronmateriaal
232 onderzoekers
241 onderzoekverloop
242 kwaliteitstoezicht
bruikbaarheid
311 doelstelling
312 toegankelijkheid
321 beantwoording vragen
322 uitvoerbaarheid lessen
slecht
matig
redelijk
goed

#	MFS-I: programmaevaluaties	#	MFS-I: programmaevaluaties
1	Local Peace Initiatives in Ituri, Congo	6	Kosovo Programma
2	The European Coalition on Oil in Sudan	7	Nagorno Karabach Programma
3	Giving the Syrian Voices of Change More Than Some Extra Oxygen!	8	Sudan Integrated Peace building Programme (SIPP)
4	The Cluster Munition Coalition	9	Norte del Cauca Programa
5	United Civilians for Peace (UCP)		

programma-evaluatiedekking	
a: mfs-I subsidiebedrag	€ 15.090.000
b: totale object programma-evaluaties	€ 3.103.876
dekking: b/a	**21 %**

ex ante kwaliteitsbeoordeling evaluatiefunctie	
(organisatie-toets subsidietoekenning 2006)	
IKV/PX-score *lerend vermogen in de keten*	8
percentage van maximum score	**89 %**

IUCN		
		MFS-I
validiteit	111 probleemstelling	
	112 evaluatiecriteria	
	121 onderzoekobject	
	122 context	
	131 interventielogica	
	132 indicatoren	
	141 analyse	
	142 conclusies	
betrouwbaarheid	211 onderzoekmethoden	
	212 triangulatie	
	221 representativiteit	
	222 beperkingen	
	231 bronmateriaal	
	232 onderzoekers	
	241 onderzoekverloop	
	242 kwaliteitstoezicht	
bruikbaarheid	311 doelstelling	
	312 toegankelijkheid	
	321 beantwoording vragen	
	322 uitvoerbaarheid lessen	

slecht	matig	redelijk	goed

MFS-i: programmaevaluatie
Ecosystems and Human Well-Being Programme

programma-evaluatiedekking	
a: mfs-I subsidiebedrag	€ 25.710.176
b: totale object programma-evaluaties	€ 15.770.361
dekking: b/a	**61 %**

ex ante kwaliteitsbeoordeling evaluatiefunctie	
(organisatie-toets subsidietoekenning 2006)	
IUCN-score *lerend vermogen in de keten*	8
percentage van maximum score	**89 %**

SOLIDARIDAD

		MFS-I		
		1	2	3
validiteit	111 probleemstelling			
	112 evaluatiecriteria			
	121 onderzoekobject			
	122 context			
	131 interventielogica			
	132 indicatoren			
	141 analyse			
	142 conclusies			
betrouwbaarheid	211 onderzoekmethoden			
	212 triangulatie			
	221 representativiteit			
	222 beperkingen			
	231 bronmateriaal			
	232 onderzoekers			
	241 onderzoekverloop			
	242 kwaliteitstoezicht			
bruikbaarheid	311 doelstelling			
	312 toegankelijkheid			
	321 beantwoording vragen			
	322 uitvoerbaarheid lessen			

slecht	matig	redelijk	goed

#	MFS-I: programmaevaluaties	#	MFS-I: programmaevaluaties
1	The Impact of Fair Trade	3	Evaluation Fashion Programme Solidaridad 2003 – 2010
2	Food, Feed and Fuel for Thought		

programma-evaluatiedekking			ex ante kwaliteitsbeoordeling evaluatiefunctie	
a: mfs-I subsidiebedrag	€ 16.619.525		(organisatie-toets subsidietoekenning 2006)	
b: totale object programma-evaluaties	€ 7.064.019		Solidaridad-score *lerend vermogen in de keten*	8
dekking: b/a	**43 %**		percentage van maximum score	**89 %**

TERRE DES HOMMES

		MFS-I			
		1	2	3	4
validiteit	111 probleemstelling				
	112 evaluatiecriteria				
	121 onderzoeksobject				
	122 context				
	131 interventielogica				
	132 indicatoren				
	141 analyse				
	142 conclusies				
betrouwbaarheid	211 onderzoeksmethoden				
	212 triangulatie				
	221 representativiteit				
	222 beperkingen				
	231 bronmateriaal				
	232 onderzoekers				
	241 onderzoek verloop				
	242 kwaliteitstoezicht				
bruikbaarheid	311 doelstelling				
	312 toegankelijkheid				
	321 beantwoording vragen				
	322 uitvoerbaarheid lessen				

slecht	matig	redelijk	goed

#	MFS-I: programmaevaluaties	#	MFS-I: programmaevaluaties
1	Pathuakali Development Project (PDP) and Coastal Island Health and Education Project (CIHEP)	3	Southeast Asia Campaign against Child Trafficking
2	Mathagondapalli Education Centre (MEC)	4	Onderwijs- en Gezondheidsprogramma's

programma-evaluatiedekking	
a: mfs-I subsidiebedrag	€ 32.090.000
b: totale object programma-evaluaties	€ 25.515.676
dekking: b/a	**80 %**

ex ante kwaliteitsbeoordeling evaluatiefunctie	
(organisatie-toets subsidietoekenning 2006)	
TdH-score *lerend vermogen in de keten*	7
percentage van maximum score	**78 %**

WOORD EN DAAD		MFS-I		
		1	2	3
validiteit	111 probleemstelling			
	112 evaluatiecriteria			
	121 onderzoekobject			
	122 context			
	131 interventielogica			
	132 indicatoren			
	141 analyse			
	142 conclusies			
betrouwbaarheid	211 onderzoekmethoden			
	212 triangulatie			
	221 representativiteit			
	222 beperkingen			
	231 bronmateriaal			
	232 onderzoekers			
	241 onderzoekverloop			
	242 kwaliteitstoezicht			
bruikbaarheid	311 doelstelling			
	312 toegankelijkheid			
	321 beantwoording vragen			
	322 uitvoerbaarheid lessen			

slecht	matig	redelijk	goed

#	MFS-I: programmaevaluaties	#	MFS-I: programmaevaluaties
1	Education Programme	3	Basic Needs Programme
2	Job and Income Programme		

programma-evaluatiedekking	
a: mfs-I subsidiebedrag	€ 19.218.872
b: totale object programma-evaluaties	€ 18.362.837
dekking: b/a	**96 %**

ex ante kwaliteitsbeoordeling evaluatiefunctie	
(organisatie-toets subsidietoekenning 2006)	
W&D-score *lerend vermogen in de keten*	9
percentage van maximum score	**100 %**

Bijlage 6 Bronnen

Adviescommissie Medefinancieringsstelsel. 2006) *Kwaliteit in ontwikkeling, Eindrapport van de Adviescommissie* MFS. Amsterdam.

Gemeenschappelijk Overleg Medefinanciering (GOM). (2003): *Medefinancieringsprogramma (MFP)-evaluatie 2003–2006*.Den Haag.

IOB (2004): *Onderzoek naar de kwaliteit van in 2002 afgeronde decentrale evaluaties*. Den Haag: IOB Evaluaties nr. 298

IOB (2005): *Onderzoek naar de kwaliteit van in 2003 afgeronde decentrale evaluaties*. Den Haag: IOB Werkdocument.

IOB. (2006). *De Kwaliteit van Programma-evaluaties in het Medefinancieringsprogramma 2003 – 2006*. Den Haag: IOB Werkdocument.

IOB. (2009). *Draagvlakonderzoek, Evalueerbaarheid en resultaten*. Den Haag: IOB Evaluaties nr. 322.

IOB. (2010). *Evaluatie van de activiteiten van de Medefinancieringsorganisaties in Nicaragua*. Den Haag: IOB Evaluaties nr.330.

MFP Breed Netwerk. (2004). *Plan van aanpak en kwaliteitscriteria tweede fase MFP-evaluaties 2003–2006*. Den Haag.

Ministerie van Financiën. (2001). *Regeling Prestatiegegevens en Evaluatieonderzoek (RPE) Rijksoverheid*. Den Haag.

Partos. (2010. *Jaarverslag 2009*. Amsterdam.

Stuurgroep Effectrapportage – Programma-Evaluatie. (1999). *Balansstudie Programma-Evaluaties 1980 – 1999*. Ede

Swanborn, P.G. (2002). *Evalueren. Het ontwerpen, begeleiden en evalueren van interventies: een methodische basis voor evaluatie-onderzoek*. Amsterdam: Boom.

Swanborn, P.G. (2005). *Methoden van sociaal-wetenschappelijk onderzoek*. Meppel: Boom.

Werkgroep IBO/MFP. (2000).*Interdepartementaal Beleidsonderzoek Medefinancieringsprogramma*. Den Haag.

Parlementaire documentatie

Brief van de minister voor Ontwikkelingssamenwerking aan de Tweede Kamer d.d. 2 oktober 2000 over de opvolging van het Interdepartementaal Beleidsonderzoek Medefinancieringsprogramma, TK: 2000–2001, 27 433, nr. 1.

Brief van de minister voor Ontwikkelingssamenwerking aan de Tweede Kamer d.d. 11 mei 2001 over en met de oorspronkelijke versie van het Beleidskader voor het Medefinancieringsprogramma-breed, TK: 2000–2001, 27 433, nr. 2.

Brief van de Staatssecretaris voor Buitenlandse Zaken aan de Tweede Kamer d.d. 18 november 2002, no. buza 020459.

Brief van de minister voor Ontwikkelingssamenwerking aan de Tweede Kamer d.d. 22 november 2001 over de herziene versie van het Beleidskader voor het Medefinancieringsprogramma-breed, TK: 2000–2001, 27 433, nr. 6.

Brief van de minister voor Ontwikkelingssamenwerking aan de Tweede Kamer d.d. 6 maart 2002 over en met het Beleidskader Thematische Medefinanciering, TK: 2001–2002, 27 433, nr. 7.

Brief van de minister voor Ontwikkelingssamenwerking aan de Tweede Kamer d.d. 3 oktober 2003 over en met de Beleidsnotitie *Aan elkaar verplicht*, *Ontwikkelingssamenwerking op weg naar 2015*, TK: 2003–2004, 29 234, nr. 1.

Brief van de minister voor Ontwikkelingssamenwerking aan de Tweede Kamer d.d. 22 december 2004 over en met het concept Beleidskader voor het Medefinancieringsstelsel, TK: 2004–2005, 27 433, nr. 29.

Brief van de minister van Financiën aan de Tweede Kamer d.d. 18 april 2006 over en met de Regeling Periodiek Evaluatieonderzoek (RPE) en Beleidsonderzoek 2006, TK; 2005–2006, 29 949 en 29 950, nr. 40.

Brief van de minister voor Ontwikkelingssamenwerking aan de Tweede Kamer d.d. 28 juni 2006 over en met het eindrapport van de evaluatie van het Thematische Medefinancieringsprogramma (TMF), TK: 2005–2006, 27 433, nr. 55.

Brief van de minister voor Ontwikkelingssamenwerking aan de Tweede Kamer d.d. 22 september 2006 over de besluitvorming met betrekking tot de subsidieronde van het Medefinancieringsstelsel (MFS) 2007–2010, TK: 2006–2007, 27 433, nr. 59.

Evaluatierapporten van de Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (IOB) gepubliceerd in 2007-2011

(Evaluatierapporten welke voor 2007 gepubliceerd zijn kunt u terugvinden op de IOB website: www.minbuza.nl/iob)

IOB nr.	Jaar	Titel evaluatierapport	ISBN
361	2011	Evaluatie van de Twinningfaciliteit Suriname-Nederland	978-90-5328-417-9
360	2011	More than Water: Impact evaluation of drinking water supply and sanitation interventions in rural Mozambique	978-90-5328-414-8
359	2011	Regionaal en geïntegreerd beleid? Evaluatie van het Nederlandse beleid met betrekking tot de Westelijke Balkan 2004-2008	978-90-5328-416-2
358	2011	Assisting Earthquake victims: Evaluation of Dutch Cooperating aid agencies (SHO) Support to Haiti in 2010	978-90-5328-413-1
357	2011	Le risque d'effets éphémères: Evaluation d'impact des programmes d'approvisionnement en eau potable et d'assainissement au Bénin	978-90-5328-415-5
357	2011	The risk of vanishing effects: Impact Evaluation of drinking water supply and sanitation programmes in rural Benin	978-90-5328-412-4
356	2011	Between High Expectations and Reality: An evaluation of budget support in Zambia	978-90-5328-411-7
355	2011	Lessons Learnt: Synthesis of literature on the impact and effectiveness of investments in education	978-90-5328-410-0
354	2011	Leren van NGOs: Studie van de basic education interventies van geselecteerde Nederlandse NGOs	978-90-5328-409-4
353	2011	Education matters: Policy review of the Dutch contribution to basic education 1999–2009	978-90-5328-408-7
352	2011	Unfinished business: making a difference in basic education. An evaluation of the impact of education policies in Zambia and the role of budget support.	978-90-5328-407-0
351	2011	Confianza sin confines: Contribución holandesa a la educación básica en Bolivia (2000-2009)	978-90-5328-406-3
350	2011	Unconditional Trust: Dutch support to basic education in Bolivia (2000-2009)	978-90-5328-405-6
349	2011	The two-pronged approach Evaluation of Netherlands Support to Formal and Non-formal Primary Education in Bangladesh, 1999-2009	978-90-5328-404-9
348	2011	Schoon schip. En dan? Evaluatie van de schuldverlichting aan de Democratische Republiek Congo 2003-2010 (Verkorte samenvatting)	978-90-5328-403-2
347	2011	Table rase – et après? Evaluation de l'Allègement de la Dette en République Démocratique du Congo 2003-2010	978-90-5328-402-5

346	2011	Vijf Jaar Top van Warschau De Nederlandse inzet voor versterking van de Raad van Europa	978-90-5328-401-8
345	2011	Wederzijdse belangen – wederzijdse voordelen Evaluatie van de Schuldverlichtingsovereenkomst van 2005 tussen de Club van Parijs en Nigeria. (Verkorte Versie)	978-90-5328-398-1
344	2011	Intérêts communs – avantages communs Evaluation de l 'accord de 2005 relatif à l 'allègement de la dette entre le Club de Paris et le Nigéria. (Version Abrégée)	978-90-5328-399-8
343	2011	Wederzijdse belangen – wederzijdse voordelen Evaluatie van de schuldverlichtingsovereenkomst van 2005 tussen de Club van Parijs en Nigeria. (Samenvatting)	978-90-5328-397-4
342	2011	Intérêts communs – avantages communs Evaluation de l'accord de 2005 relatif à l'allègement de la dette entre le Club de Paris et le Nigéria. (Sommaire)	978-90-5328-395-0
341	2011	Mutual Interests – mutual benefits Evaluation of the 2005 debt relief agreement between the Paris Club and Nigeria. (Summary report)	978-90-5328-394-3
340	2011	Mutual Interests – mutual benefits Evaluation of the 2005 debt relief agreement between the Paris Club and Nigeria. (Main report)	978-90-5328-393-6
338	2011	Consulaire Dienstverlening Doorgelicht 2007-2010	978-90-5328-400-1
337	2011	Evaluación de las actividades de las organizaciones holandesas de cofinanciamiento activas en Nicaragua	-
336	2011	Facilitating Resourcefulness. Synthesis report of the Evaluation of Dutch support to Capacity Development.	978-90-5328-392-9
335	2011	Evaluation of Dutch support to Capacity Development. The case of the Netherlands Commission for Environmental Assessment (NCEA)	978-90-5328-391-2
	2011	Aiding the Peace. A Multi-Donor Evaluation of Support to Conflict Prevention and Peacebuilding Activities in Southern Sudan 2005 - 2010	978-90-5328-389-9
333	2011	Evaluación de la cooperación holandesa con Nicaragua 2005-2008	978-90-5328-390-5
332	2011	Evaluation of Dutch support to Capacity Development. The case of PSO	978-90-5328-388-2
331	2011	Evaluation of Dutch support to Capacity Development. The case of the Netherlands Institute for Multiparty Democracy (NIMD)	978-90-5328-387-5
330	2010	Evaluatie van de activiteiten van de medefinancieringsorganisaties in Nicaragua	978-90-5328-386-8
329	2010	Evaluation of General Budget Support to Nicaragua 2005-2008	978-90-5328-385-1
328	2010	Evaluatie van de Nederlandse hulp aan Nicaragua 2005-2008	978-90-5328-384-4

327	2010	Impact Evaluation. Drinking water supply and sanitation programme supported by the Netherlands in Fayoum Governorate, Arab Republic of Egypt, 1990-2009	978-90-5328-381-3
326	2009	Evaluatie van de Atlantische Commissie (2006-2009)	978-90-5328-380-6
325	2009	Beleidsdoorlichting van het Nederlandse exportcontrole- en wapenexportbeleid	978-90-5328-379-0
-	2009	Evaluatiebeleid en richtlijnen voor evaluaties	-
324	2009	Investing in Infrastructure	978-90-5328-378-3
-	2009	Synthesis of impact evaluations in sexual and reproductive health and rights	978-90-5328-376-9
323	2009	Preparing the ground for a safer World	978-90-5328-377-6
322	2009	Draagvlakonderzoek. Evalueerbaarheid en resultaten	978-90-5328-375-2
321	2009	Maatgesneden Monitoring 'Het verhaal achter de cijfers'	978-90-5328-374-5
320	2008	Het tropisch regenwoud in het OS-beleid 1999-2005	978-90-5328-373-8
319	2008	Meer dan een dak. Evaluatie van het Nederlands beleid voor stedelijke armoedebestrijding	978-90-5328-365-3
318	2008	Samenwerking met Clingendael	978-90-5328-367-7
317	2008	Sectorsteun in milieu en water	978-90-5328-369-1
316	2008	Be our guests (sommaire)	978-90-5328-372-1
316	2008	Be our guests (summary)	978-90-5328-371-4
316	2008	Be our guests (hoofdrapport Engels)	978-90-5328-371-4
316	2008	Be our guests (samenvatting)	978-90-5328-370-7
316	2008	Be our guests (hoofdrapport)	978-90-5328-370-7
315	2008	Support to Rural Water Supply and Sanitation in Dhamar and Hodeidah Governorates, Republic of Yemen	978-90-5328-368-4
314	2008	Primus Inter Pares; een evaluatie van het Nederlandse EU-voorzitterschap 2004	978-90-5328-364-6
313	2008	Explore-programma	978-90-5328-362-2
312	2008	Impact Evaluation: Primary Education Zambia	978-90-5328-360-8
311	2008	Impact Evaluation: Primary Education Uganda	978-90-5328-361-5
310	2008	Clean and Sustainable?	978-90-5328-356-1
309	2008	Het vakbondsmedefinancieringsprogramma – samenvatting Engels	978-90-5328-357-8
309	2008	Het vakbondsmedefinancieringsprogramma – Samenvatting Spaans	978-90-5328-357-8
309	2008	Het vakbondsmedefinancieringsprogramma	978-90-5328-357-8
308	2008	Het Nederlandse Afrikabeleid 1998-2006. Evaluatie van de bilaterale samenwerking	978-90-5328-359-2
308	2008	Het Nederlandse Afrikabeleid 1998-2006. Evaluatie van de bilaterale samenwerking (Samenvatting)	978-90-5328-359-2

307	2008	Beleidsdoorlichting seksuele en reproductieve gezondheid en rechten en hiv/aids 2004-2006	978-90-5328-358-5
306	2007	Chatting and Playing Chess with Policymakers	978-90-5328-355-4
305	2007	Impact Evaluation: Water Supply and Sanitation Programmes Shinyanga Region, Tanzania 1990-2006	978-90-5328-354-7
304	2007	Evaluatie van de vernieuwing van het Nederlandse onderzoeksbeleid 1992-2005	978-90-5328-353-0
304	2007	Evaluation of the Netherlands Research Policy 1992-2005 (Summary)	978-90-5328-353-0

Uitgebracht door:

Ministerie van Buitenlandse Zaken
Postbus 20061 | 2500 EB Den Haag
www.minbuza.nl | www.rijksoverheid.nl

Opmaak: Vijfkeerblauw Rijswijk
Druk: OBT bv, Den Haag
ISBN: 978-90-5328-418-6

© Ministerie van Buitenlandse Zaken | December 2011