

juni 2006

de
Kwaliteit
van
**Programma-
evaluaties**
in het
**Medefinancierings-
programma**
2003 – 2006

Dick van der Hoek



**Buitenlandse
Zaken**

de
Kwaliteit
van
Programma-evaluaties
in het
**Medefinancierings-
programma**
2003 – 2006

Dick van der Hoek

VOORWOORD

Dit werkdocument is het resultaat van een *meta-evaluatie*: een evaluatie van evaluaties. Het doel ervan was na te gaan in hoeverre de programma-evaluatie-onderzoeken die vijf van de zes medefinancieringsorganisaties tijdens de subsidieperiode van het Medefinancieringsprogramma-breed (2003 t/m 2006) hebben uitgevoerd, geschikt zijn voor de verantwoording van beleid en uitvoering. Om die geschiktheid te bepalen is – evenals bij eerder uitgevoerde beoordelingen van bilaterale evaluatiestudies – gebruik gemaakt van kwaliteitseisen die werden ontleend aan de Regeling Prestatiegegevens en Evaluatieonderzoek Rijksoverheid.

Het rapport is bestemd voor de Directie Sociale en Institutionele Ontwikkeling van dit ministerie ten behoeve van het beheer van het MFP-breed, waarmee deze directie is belast. Daarnaast bestaat uiteraard de hoop dat de betrokken MFO's de bevindingen van het onderzoek ter harte zullen nemen, om er vervolgens hun voordeel mee te doen bij de verbetering van hun toekomstige evaluatie-inspanningen.

Het onderzoek is verricht door IOB-evaluator Dick van der Hoek, maar had niet kunnen worden uitgevoerd zonder de medewerking van de evaluatievertegenwoordigers van Cordaid, Hivos, Icco, Novib en Plan Nederland, waarvoor hun allen dank is verschuldigd. Zij dragen echter geen medeverantwoordelijkheid voor de inhoud van het rapport. Die berust geheel bij deze inspectie.

Henri Jorritsma

Waarnemend Directeur Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie

INHOUD

<i>Voorwoord</i>	<i>iii</i>	
<i>Tabellen en figuren</i>	<i>vi</i>	
<i>Afkortingen</i>	<i>vii</i>	
1	HOOFDBEVINDINGEN	1
2	INLEIDING	3
2.1	Aanleiding	3
2.2	Doelstelling	4
2.3	Probleemstelling	4
2.4	Rapportopzet	4
3	ONDERZOEKSOBJECT	5
3.1	Projectevaluaties	5
3.2	Programma-evaluaties	6
3.3	Afbakening	9
4	ONDERZOEKSAANPAK	11
4.1	Beoordelingslijst	11
4.2	Waarderingsmethode	12
4.3	Beoordelingsprocedure	14
5	ONDERZOEKSRISULTATEN	17
5.1	Algemene kenmerken	17
5.2	Validiteit	20
5.3	Betrouwbaarheid	23
5.4	Bruikbaarheid	25
	BIJLAGEN	29
1	Kenmerken MBN Programma-evaluaties	31
2	Kenmerken Individuele Programma-evaluaties	33
3	Beoordelingslijst	45
4	Beoordelingscores	51

TABELLEN EN FIGUREN

TABELLEN

3-1	Aantal Individuele Programma-evaluaties per MFO 2003-2006	8
5-1	Aanleiding en doelstelling	20
5-2	Probleemstelling en onderzoeksvragen	21
5-3	Evaluatie-object en context	21
5-4	Interventiologica	22
5-5	Onderzoeksmethoden	22
5-6	Consistentie en soliditeit	23
5-7	Gegevensbronnen	23
5-8	Representativiteit	24
5-9	Onafhankelijkheid	24
5-10	Onderzoeksverloop	25
5-11	Kwaliteitsbewaking	25
5-12	Aansluiting op behoeften	26
5-13	Volledigheid	26
5-14	Presentatie	26

FIGUREN

4-1	Beoordelingspiramide	11
4-2	Aanlevering PE-rapporten per kwartaal 2003-2006	14

AFKORTINGEN

AKV	Apparaatskostenvergoeding
DSI	Directie Sociale en Institutionele Ontwikkeling
E&B	Evaluatie & Beleid (MBN-werkgroep)
ERG	Evaluation Reference Group
fte	full-time equivalent (volledige formatieplaats)
GOM	Gemeenschappelijk Overleg Medefinanciering
IBO	Interdepartementaal Beleidsonderzoek
IOB	Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie
MBN	Medefinancieringsprogramma Breed Netwerk
M&E	Monitoring en Evaluatie
MFO	Medefinancieringsorganisatie
MFP	Medefinancieringsprogramma
OS	Ontwikkelingssamenwerking
PE	Programma-evaluatie
PO	Partnerorganisatie
RPE	Regeling Prestatiegegevens en Evaluatieonderzoek
<i>SMART</i>	<i>Specific – Measurable – Attainable –Relevant –Time-bound</i>
ToR	Terms of Reference
VBTB	Van Beleidsbegroting Tot Beleidsverantwoording

1 HOOFDBEVINDINGEN

Inleiding

Bij de instelling van het Medefinancieringsprogramma Breed werd bepaald dat de Inspectie Ontwikkelingssamenwerking en Beleidsevaluatie (IOB) toezicht diende te houden op de kwaliteit van evaluaties die de Medefinancieringsorganisaties (MFO's) zelf uitvoeren. Daartoe beoordeelde IOB alle 26 programma-evaluaties (PE's) die tijdens de subsidieperiode van het MFP-breed: 2003 – 2006 zijn uitgevoerd. De gegevensverzameling werd afgesloten op 1 maart 2006. PE's zijn over het algemeen breder van opzet dan evaluaties van individuele projecten en beogen conclusies te trekken op een beleidsniveau dat dat van de individuele samenwerkingsactiviteit overstijgt. Vier van de beoordeelde studies maakten deel uit van de meerjaren onderzoeksagenda van het MFP-breed Netwerk (MBN). De overige 22 hadden betrekking op evaluatiestudies ondernomen door individuele MFO's. Met de activiteiten die in de 26 onderzoeken zijn geëvalueerd, waren programma-uitgaven ten bedrage van € 364 mln gemeoid. De bestudeerde onderzoeken kostten in totaal € 1,9 mln (exclusief MFO apparaatskosten).

De beoordeling is uitgevoerd aan de hand van de drie kwaliteitscriteria ontleend aan de Regeling Prestatiegegevens en Evaluatie-onderzoek (RPE) voor de Rijksoverheid: Validiteit, Betrouwbaarheid en Bruikbaarheid. Deze criteria zijn geoperationaliseerd door de identificatie van 14 indicatoren, die vervolgens verder werden geconcretiseerd door ze uit te splitsen in de 32 componenten. De 26 PE's zijn op elk van deze 32 componenten beoordeeld, waarbij de scores werden vastgelegd op een 4-puntsschaal met waarderingen variërend van *slecht* via *matig* en *redelijk* tot *goed*. De beoordelingen zijn tot stand gekomen na overleg tussen IOB enerzijds en betrokken MFO-evaluatievertegenwoordigers en -programmaverantwoordelijken anderzijds in twee rondes: een schriftelijke en een mondelinge. Tijdens de eerste (schriftelijke) ronde werd vooral aanvullende informatie opgevraagd en verstrekt, terwijl in de tweede de toelichting op conceptwaarderingen en de discussie daarover centraal stond. Na eventuele aanpassing van conceptscores werden de PE-beoordelingen ten slotte vastgesteld.

Algemeen

Zoals uit de scoreverdelingen per beoordelingscomponent blijkt, is voor tweederde van deze componenten de waardering overwegend positief, dat wil zeggen dat de score *goed* of *redelijk* is. Daarentegen vertonen de bestudeerde onderzoeken op essentiële onderdelen – vermeld bij de drie RPE kwaliteitscriteria hieronder – ook een aantal tekortkomingen. Dit zijn er te veel om, over het geheel genomen, programma-evaluaties in hun huidige uitvoering te kunnen kwalificeren als een adequaat instrument voor de verantwoording van bestedingen.

Weliswaar scoren de MBN-evaluaties door de bank genomen iets beter dan de individuele PE's, maar die verschillen zijn niet groot genoeg om voor de gezamenlijke onderzoeken een andere conclusie te rechtvaardigen.

Validiteit

Uit de waarderingen op componenten als aanleidingen, doelstellingen en probleemstellingen van evaluatie-onderzoeken blijkt dat de opdrachtgevers van PE's doorgaans goed weten waar ze willen uitkomen en waarom ze dat willen.

Het gaat echter vaak mis op het moment dat die bedoelingen moeten worden uitgewerkt in concrete onderzoeksontwerpen. Vooral bij de reconstructie van de interventielogica, de identificatie van resultaatniveaus en indicatoren, en de definiëring van gehanteerde criteria blijft veel te wensen over. In mindere mate geldt dat voor onderdelen als de gegevensanalyse en de

formulering van onderzoeksvragen. De meest winstgevende mogelijkheden tot verbetering bevinden zich daarmee bij de onderdelen die bepalend zijn voor de transparantie van het evaluatie-onderzoek.

Betrouwbaarheid

De meest positieve waarderungen onder dit criterium gelden de onafhankelijkheid van de ingeschakelde onderzoekers, de inspraak van belanghebbenden via becommentariëring en de identificatie van gebruikte menselijke en documentaire gegevensbronnen.

De betrouwbaarheid van PE's wordt volgens dit onderzoek aangetast door het geringe gebruik van bronnen die onafhankelijk zijn van de belanghebbenden en van het object van de evaluatie en door de beperkte onafhankelijkheid van toezicht en begeleiding bij de individuele (interne) PE's. Het belangrijkste obstakel voor voldoende betrouwbaarheid wordt echter gevormd door de bijna stelselmatig tekortschietende aandacht voor de representativiteit van geselecteerde case studies en bijgevolg voor de generaliseerbaarheid van de evaluatiebevindingen.

Bruikbaarheid

Zoals uit de beoordelingstabellen valt af te leiden is de bruikbaarheid, die met name betrekking heeft op de eindrapportage, zeker voor de directe betrokkenen bij de MFO's die de resultaten voornamelijk nodig hebben voor nieuwe beleidsontwikkeling, alleszins redelijk.

De bruikbaarheid voor verantwoordingsdoeleinden wordt onder dit criterium echter het meest weerspiegeld in de waardering voor de verklaring van verschillen tussen beoogde en behaalde resultaten. De lage scores op dit onderdeel hangen uiteraard samen met het feit dat in de meeste PE's de reconstructie van de interventielogica tekortschiet of geheel ontbreekt. Waarmee onderstreept wordt dat de bruikbaarheid van evaluatie-uitkomsten in ieder geval voor een deel afhangt van de validiteit van het onderzoek waar zij uit voortkomen.

2 INLEIDING

Dit hoofdstuk beoogt in kort bestek de achtergrond van het IOB-onderzoek naar de kwaliteit van programma-evaluaties¹ van de Medefinancieringsorganisaties (MFO's) te schetsen werd opgezet en ten uitvoer gelegd.

2.1 Aanleiding

Met de invoering van het programmafinancieringsmodel in 1980 deed ook de zogenoemde programma-evaluatie (PE) haar intrede in het Medefinancieringsprogramma (MFP). Deze PE's betroffen doorgaans onderzoeken rond een bepaald thema of in een specifieke regio, waarbij de activiteiten van één MFO centraal stonden. Tussen 1980 en 1999 werden er in opdracht van het ministerie van Buitenlandse Zaken en het Gemeenschappelijk Overleg Medefinanciering (GOM)² zo'n 80 van dergelijke studies uitgevoerd. Aan de uitvoering ervan namen de MFO's zelf actief deel, hetgeen tegen het einde van de jaren negentig leidde tot toenemende kritiek op het gebrek aan onafhankelijkheid van de onderzoeken. Bovendien zou de PE-rapportage in het algemeen onvoldoende resultaatgericht zijn, althans te weinig informatie bevatten over behaalde resultaten.

Om aan deze bezwaren tegemoet te komen werd in 1998 overgegaan tot instelling van de onafhankelijke Stuurgroep Evaluatie Medefinancieringsprogramma, belast met de opzet en uitvoering van de programma-evaluaties. 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 programma-evaluaties. Het resultaat was de *Balansstudie* die onder meer een 'systematisch falen van de effectmetingsfunctie van PE's' vaststelde.³

Aan deze laatste bevinding werd groot gewicht toegekend door het Interdepartementaal Beleids-onderzoek (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 onvoldoende inzicht [kon] worden verworven in de doeltreffendheid en doelmatigheid van het MFP en dat de mogelijkheid tot aansturing van het programma door de minister voor Ontwikkelingssamenwerking tekort [schoot].⁴ Behalve een duidelijker beleidskader bepleitte het IBO ook een scherper toezicht op de naleving ervan. In overeenstemming daarmee formuleerde de minister voor OS een geheel nieuw raamwerk voor het medefinancieringsprogramma: het Beleidskader MFP-breed. In de definitieve versie daarvan die per 1 januari 2003 voor een periode van vier jaar van kracht werd, is de rol van IOB 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.*⁵

¹ Hoewel programmaevaluatie eigenlijk één woord is, wordt het ter wille van de leesbaarheid doorgaans met een koppelteken gespeld: programma-evaluatie. Wellicht hiermee samenhangend is de in het MFP gangbare afkorting: PE.

² In 2004 omgedoopt tot het MFP Breed Netwerk (MBN).

³ Balansstudie (1999): p. 45.

⁴ Volgens de brief waarmee de minister voor OS op 2 oktober 2000 het IBO-rapport aan de Tweede Kamer aanbod, TK: 2000–2001, 27 433, nr. 1, p. 1.

⁵ Aangepaste versie van het nieuwe 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.

Aan de invulling van dit mandaat wordt op twee manieren vorm gegeven:

- Door middel van een beperkt aantal 'eigen' IOB-evaluaties. Dit zijn zelfstandige Inspectie-onderzoeken naar werkwijzen en resultaten van MFO-activiteiten, uitmondend in afzonderlijke evaluatierapporten. Zij maken geen deel uit van dit onderzoek.
- Door toezicht op de kwaliteit van de evaluaties die de MFO's zelf uitvoeren.⁶

Het voorliggende rapport bevat de resultaten van dit toezicht op de kwaliteit van de programma-evaluaties binnen het MFP-breed gedurende de subsidieperiode 2003–2006.

2.2 Doelstelling

Tegen deze achtergrond van de geschetste voorgeschiedenis ligt het voor de hand dat dit onderzoek in de eerste plaats een verantwoordingsfunctie heeft. Daarmee is dat de primaire doelstelling.

Uiteraard hoopt de Inspectie dat de bevindingen en conclusies van deze studie een bijdrage kunnen leveren aan de kwaliteitsverbetering van toekomstig evaluatie-onderzoek. In de mate waarin dat het geval is, krijgt deze evaluatie dan naast haar primaire verantwoordingsfunctie tevens een leerfunctie. Of die laatste daadwerkelijk uit de verf komt, hangt grotendeels af van de betrokken MFO's en van hun opstelling ten opzichte van de onderzoeksresultaten.

2.3 Probleemstelling

Geeft de onderzoeksdoelstelling aan waarvoor men de uitkomsten wil gebruiken, bij de probleemstelling gaat het erom welke centrale onderzoeksvraag of probleemstelling daartoe moet worden beantwoord. In het onderhavige geval luidt die als volgt:

In welke mate voldoen programma-evaluaties binnen het Medefinancieringsprogramma-breed 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 *Regeling Prestatiegegevens en Evaluatieonderzoek (RPE)* van de Rijksoverheid, waarmee de drie onderzoeksvragen worden (i) *hoe valide*, (ii) *hoe betrouwbaar* en (iii) *hoe bruikbaar* de onderzochte PE's zijn. Verdere concretisering van de onderzoeksvragen vindt plaats in hoofdstuk 4 over de onderzoeksaanpak.

2.4 Rapportopzet

Na deze inleiding wordt in het derde hoofdstuk het object van deze studie geïntroduceerd: de verzameling van alle gezamenlijke en individuele programma-evaluaties die de MFO's gedurende de lopende subsidieperiode hebben ondernomen. Hoofdstuk 4 zet daarna uiteen hoe dat onderzoek is aangepakt, waarna hoofdstuk 5 tenslotte weergeeft tot welke uitkomsten dat heeft geleid. De belangrijkste daarvan zijn aan het begin van dit rapport samengevat in het eerste hoofdstuk onder hoofdbevindingen en aandachtspunten.

⁶ Brief van de Staatssecretaris voor Buitenlandse zaken aan de Tweede Kamer no. buza 020459 van 18 november 2002, p. 4.

3 ONDERZOEKSOBJEKT

Het toezicht waarvoor de Inspectie in 2002 opdracht ontving, had betrekking op de kwaliteit van wat in de RPE regeling wordt aangeduid als *evaluatie-onderzoek ex post*. Volgens definitiebepaling 2.2 van deze regeling 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 MFP ondernomen studies aan deze opzet zouden kunnen beantwoorden vond kort na de start van de subsidieperiode (5 maart 2003) overleg plaats tussen de evaluatievertegenwoordigers van alle zes MFO's en IOB. Hierbij bleek het nuttig onderscheid te maken tussen drie soorten of aggregatieniveaus van evaluatie:

- Projectevaluaties
- Individuele programma-evaluaties van verscheidene vergelijkbare interventies van één MFO.
- Gezamenlijke programma-evaluaties (PE's) van vergelijkbare activiteiten van verscheidene MFO's, in het kader van het Gemeenschappelijk Overleg Medefinanciering (GOM).

3.1 Projectevaluaties

Onder *projectevaluaties* worden de doorgaans kleinschalige evaluatie-onderzoeken verstaan die zich beperken tot individuele projecten, financieringsovereenkomsten en/of relaties met individuele partnerorganisaties (PO's). 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 betreffende MFO – 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 het object van onderzoek vormen. In sommige gevallen draagt de PO zelfs formeel de verantwoordelijkheid 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 MFO 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 evaluatie-rapport, maar om het *hele* onderzoek dat eraan ten grondslag heeft gelegen. Gezien de vaak dominante invloed die het evaluatie-ontwerp en de sturing van het evaluatieproces op de uiteindelijke onderzoekskwaliteit hebben, is het wenselijk aan die onderdelen bijzondere aandacht te besteden. En dat vergt doorgaans toegang tot relevante menselijk bronnen, zoals programmabeheerders en personeel van evaluatie-eenheden.

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 evaluatie-onderzoek hebben soms de neiging het onderscheid tussen beide functies kunstmatig zwaar aan te zetten, alsof zij

elkaar zouden uitsluiten. Langdurige ervaring met Inspectie-onderzoeken wijst echter uit dat dit niet het geval is en dat beide functies met goed opgezet, resultaatgericht evaluatie-onderzoek zeer wel kunnen worden gediend. Wel richten de leer- en verantwoordingsfuncties zich meestal op verschillende doelgroepen. Waar het afleggen van verantwoording vooral wordt geassocieerd met categorieën van belanghebbenden als opdrachtgevers en donateurs, bestaat de primaire doelgroep van de leerfunctie uit vormgevers, uitvoerders en begunstigden van de geëvalueerde activiteiten.⁷ Een belangrijk gevolg van dit verschil is dat zogenaamde *formatieve* evaluaties kunnen volstaan met een beperkte hoeveelheid feitelijke en achtergrondinformatie over evaluatieobject en –context, omdat de beoogde gebruikersgroep van de onderzoeksrapportage bestaat uit ingewijden. Voor buitenstaanders zijn dergelijke onderzoeksrapporten veelal ontoegankelijk, omdat ze te veel bekend veronderstellen. Verder zijn ‘leerevaluaties’ doorgaans sterk op de toekomst gericht, omdat niet zozeer het voortbestaan van het evaluatie-object staat ter discussie, 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 (concept-) beleid dan op de empirische vaststelling van bereikte resultaten.

Dit effect wordt ten slotte vaak nog versterkt doordat voor de betrokken MFO projectevaluaties een belangrijke rol spelen bij de voorbereiding van managementbeslissingen over de verlenging of verbreking van contracten, relaties met partnerorganisaties, personele inzetten en dergelijke. Deze beheersfunctie van projectevaluaties bevordert welhaast onvermijdelijk de reeds aanwezige toekomstgerichtheid van de betreffende onderzoeken en rapportages ten koste van het empirische karakter.

3.2 Programma-evaluaties (PE's)

Programma-evaluaties verschillen van projectevaluaties doordat ze een bredere opzet hebben en een omvangrijker onderzoeksobject bestrijken dan het individuele project of de relatie met één enkele PO. Ze kunnen worden onderscheiden in *gezamenlijke* en *individuele* PE's. De gezamenlijke onderzoeken worden door meerdere MFO's ondernomen in het kader van hun samenwerking binnen het MFP-breed Netwerk en daarom ook wel aangeduid als MBN programma-evaluaties. Bij een individuele PE, soms interne PE genoemd, gaat het om één afzonderlijke MFO die een reeks van haar eigen activiteiten evalueert of laat evalueren.

3.2.1 Individuele PE's

Tijdens het overleg van IOB met de MFO's in maart 2003 over de invulling van de kwaliteitsbeoordeling van hun evaluatie-onderzoek was niet bekend hoeveel individuele programma-evaluaties er jaarlijks werden ondernomen. De veronderstelling was dat het aantal zou liggen tussen de omvang van de geplande serie MBN-PE's (die destijds zeven onderzoeken omvatte) en de honderden projectevaluaties waarvan de MFO's samen in hun jaarverslagen melding maakten. Later bleek dat de onzekerheid over het aantal te verwachten PE's vooral de onduidelijkheid over de inhoud van het *begrip* programma-evaluatie weerspiegelde. De verwarring werd zichtbaar toen individuele MFO's in de loop van 2004 de meest uiteenlopende onderzoeksrapporten gingen aanleveren als 'programma-evaluatie'. Een aantal bleek zich niet te lenen voor toetsing in het kader van de IOB-toezicht (zie volgende hoofdstuk), omdat ze gewoonweg niet als evaluatierapporten konden worden beschouwd.

⁷ Naar aanleiding van de bespreking van 5 maart 2003 merkte één van de MFO-evaluatievertegenwoordigers tegenover IOB op: "U kunt zich voorstellen dat er zich een ongewenste situatie kan voordoen wanneer projectevaluaties met als expliciete hoofddoelstelling "learning" [...] gebruikt zouden worden in een "accountability"-exercitie."

Een en ander maakte duidelijk dat het nodig was een heldere en hanteerbare omschrijving van het begrip 'programma-evaluatie' te ontwikkelen. Hierbij is een pragmatische benadering gevolgd door alle aangeleverde rapporten te bestuderen en vervolgens te controleren op hun toetsbaarheid aan de evaluatiekwaliteitscriteria uit de RPE. Over die toetsbaarheid werd, met name in twijfelgevallen, uiteraard overleg gepleegd met betrokken MFO-functionarissen. Hieruit ontstonden geleidelijk enkele praktische inzichten die volgens alle betrokkenen een hanteerbare afbakening van het begrip programma-evaluatie opleverden, waarbij met name drie dimensies van belang bleken:

- De term **programma** blijkt in de praktijk te worden gebruikt als een containerbegrip, waarin allerlei soorten verzamelingen passen. Bijgevolg dient een PE te worden gezien als een evaluatie waarvan het onderzoeksobject het niveau van het individuele project, de individuele financiering en de individuele MFO-PO relatie overstijgt. Dit houdt in dat een PE-onderzoek in ieder geval betrekking moet hebben op een *reeks* van activiteiten met als gemeenschappelijk kenmerk een *sector* (bijv. onderwijs), een *dwarsdoorsnijdend thema* (bijv. gender), een *land* (bijv. Bolivia), een *regio binnen een land* (bijv. Noord-Irak), *landen binnen een regio* (bijv. Centraal-Azië) of een *interventiestrategie* (bijv. beleidsbeïnvloeding). Daarbij doet het er niet toe of de betreffende MFO's de geëvalueerde reeks activiteiten ook buiten het evaluatie-onderzoek een *programma* noemen.
- Een PE is gericht op **resultaatmeting**. Die resultaten kunnen op *output*, *outcome* of *impact*-niveau liggen, maar zijn in ieder geval duidelijk te onderscheiden van allerlei procesvariabelen als intern-organisatorische maatregelen, klantentevredenheidsscores, (back)donor-relaties, enz. Gezien vanuit de RPE die de basis vormt voor de IOB-kwaliteitsbeoordeling, bevinden dergelijke 'procesresultaten' zich aan de *input*-kant van de interventielogica.
- Het mandaat van IOB betreft de besteding van **MFP-fondsen** uit de OS-begroting. Dit betekent dat, om voor een beoordeling in aanmerking te komen, de in een PE onderzochte activiteiten tenminste gedeeltelijk met MFP-middelen dienen te zijn gefinancierd. De omvang van het beslag op de MFP-begroting is een voor de beoordeling relevant kenmerk (zie Beoordelingslijst in bijlage 3).

Ook vanuit het perspectief van het gebruik van de uitkomsten van evaluatie-onderzoek kan een onderscheid 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 teruggekoppeld naar sectoraal, thematisch, regionaal of strategisch *beleidsniveau*.

Halverwege de subsidieperiode, begin maart 2005, werd door IOB een inventarisatie gemaakt van de tot dan toe opgedane ervaringen en vervolgens met de Werkgroep Evaluatie & Beleid (E&B) van het MBN besproken. Bij die gelegenheid (21 juni 2005) werd de afbakening van het PE-begrip in de hiervoor beschreven zin vastgelegd. Afgaande op de sterke daling van de sinds die tijd ontvangen ongeschikte onderzoeksrapporten is de beoogde verduidelijking gerealiseerd.

Bij de afsluiting van de gegevensverzameling op 1 maart 2006 waren 25 van de 33 ontvangen onderzoeksrapporten als programma-evaluatie bestempeld en dienovereenkomstig beoordeeld. De beoordeelde individuele PE's waren als volgt over de vijf deelnemende MFO's verdeeld:

Tabel 3–1 Aantal Individuele Programma-evaluaties per MFO 2003–2006

MFO ⁸ :	Cordaid	Hivos	Icco	Novib	Plan
Aantal PE's:	5	5	5	7	3

Bron: IOB archief MFP PE kwaliteit

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. In plaats van 25 bevat deze bijlage slechts 22 onderzoeksrapporten, omdat in drie gevallen de betreffende studies coproducties vormden van Cordaid en Icco (2x) en van Hivos en Novib (1x) ter evaluatie van gezamenlijk gefinancierde en uitgevoerde activiteiten.

3.2.2 MBN PE's

Met de beëindiging van het mandaat van de onafhankelijk Stuurgroep Evaluatie Medefinancierings-programma, die in de periode 1999 – 2002 belast was met de evaluatie van het MFP, 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 evaluatie-onderzoek.

Daartoe stelde de E&B Werkgroep van het GOM⁹ eind 2002 een gezamenlijke meerjaren onderzoeksagenda op. Deze gold voor de hele subsidieperiode en bevatte zeven hoofdonderwerpen voor evaluatie-onderzoek, 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 niveau van bereik en outcome op doelgroepniveau) en hoe zijn deze te beoordelen?*¹⁰ Voor elk onderzoek zou een zogenaamde Coördinatiegroep worden geformeerd met vertegenwoordigers van de deelnemende MFO's en één van de GOM-directeuren als voorzitter.

Van de zeven geplande onderzoeken zijn er vijf uitgevoerd, terwijl twee thema's werden geschrapt: *Vrede, veiligheid en conflict* en *Duurzame Economische Ontwikkeling (DEO)*. Het laatste onderwerp is in 2005 door IOB geselecteerd als een van de terreinen voor een 'eigen evaluatie' waarnaar in het hiervoor besproken IOB-mandaat (§ 2.1) wordt verwezen. De eerste van de vijf ondernomen evaluatiestudies (*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 leerdokument".

Dit laatste gebeurde mede op advies van een inmiddels ingestelde *Evaluation Reference Group (ERG)*¹¹ bestaande uit een zevental onafhankelijke, externe deskundigen met als mandaat de

⁸ Terre des Hommes kende tot de toetreding tot het MFP alleen evaluatie-onderzoek op project- of counterpartniveau en is eerst in de loop van de subsidieperiode 2003–2006 begonnen met de introductie van breder opgezette programma-evaluaties. Bij de afsluiting van de gegevensverzameling voor deze studie waren hiervan nog geen resultaten beschikbaar.

⁹ Vijf van de zes MFO's (Cordaid, Hivos, ICCO, Novib en Plan-Nederland) waren lid van het GOM (en zijn het dus nu van het MBN). De zesde, Terre des Hommes, trad niet toe tot het samenwerkingsverband en nam dus ook geen deel aan de gezamenlijke programma-evaluaties.

¹⁰ 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 partnerorganisaties 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*).

¹¹ Soms wordt deze afkorting uitgelegd als staande voor *External Reference Group*. Beide betekenissen sporen overigens met de werkelijkheid: de Referentiegroep is zowel *extern* t.o.v. het MBN als ingesteld t.b.v. *evaluaties*.

beoordeling van de kwaliteit van proces en resultaten van de MBN evaluatiestudies. Hiertoe zou men zich richten op vier gebieden: de overkoepelende *Terms of Reference (ToR)*, het evaluatie-ontwerp, het eindproduct en het programma-evaluatieproces. De ERG begeleidde en becommentarieerde na het onderzoek *Basiseducatie* eveneens de vier nog resterende MBN-studies die wel zijn uitgemond in volwaardige PE-rapporten. De voltooide onderzoeken, waarvan de belangrijkste kenmerken zijn opgenomen in bijlage 1, betroffen achtereenvolgens de onderwerpen: (i) Rol van Vrouwenorganisaties in maatschappijopbouw, (ii) Gezondheidszorg, (iii) Civil Society Participation en (iv) HIV/AIDS.

3.3 Afbakening

Dienen projectevaluaties vooral de leer- en beheersfuncties van evaluatie-onderzoek, bij programma-evaluaties nemen, volgens de MFO's zelf,¹² de vaststelling en verantwoording van resultaten op *outcome* en *impact*niveau een centrale plaats bij onderzoeksopzet en –uitvoering. Dit gegeven leidt in combinatie met hiervoor aangegeven overwegingen van praktische aard, tot de beslissing om projectevaluaties van dit onderzoek uit te zonderen. Bijgevolg zal de kwaliteitsbeoordeling zich richten op de 4 gezamenlijke MBN programma-evaluaties en de 22 individuele PE's van Cordaid, Hivos, Icco, Novib en Plan die in de lopende subsidieperiode tot 1 maart 2006 werden afgerond en bij IOB aangemeld. Omdat Terre des Hommes geen deel uitmaakt van het MBN en ook geen individueel PE-onderzoek uitvoerde, blijft deze MFO voor de onderhavige studie buiten beschouwing.

3.3.1 Beperkingen

Van de traditionele functies waarvoor evaluatiestudies kunnen worden gebruikt: (i) verantwoorden, (ii) leren en (iii) beheren, concentreert dit onderzoek zich geheel op de eerste. Dat betekent dat de kwaliteit van evaluatie-onderzoek afhangt van zijn geschiktheid als instrument om verantwoording af te leggen over gevoerd beleid en verrichte uitgaven. Daarmee is niet gezegd of gesuggereerd dat de andere gebruiksdoelen van evaluatie minder belangrijk zouden zijn. De concentratie op verantwoording vloeit eenvoudigweg voort uit het feit dat het analysekader dat IOB dient te hanteren bij de uitvoering van haar mandaat ten opzichte van het MFP-breed, wordt gevormd door de RPE. 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)*.

MFO's leggen niet alleen verantwoording af door middel van evaluatie-onderzoek, maar doen dat ook op andere manieren, bijvoorbeeld via (verplicht) periodiek accountantsonderzoek, dat zich in het bijzonder richt op de rechtmatigheid van de verrichte uitgaven. Dergelijke verantwoordingsinstrumenten blijven in de onderhavige studie buiten beschouwing, evenals onderzoeken naar klantentevredenheid, omdat die vanuit de RPE beschouwd slechts een tussen- of procesresultaat vormt.

¹² Zoals onder meer gesteld in brieven aan IOB van de Voorzitter van de MBN werkgroep E&B d.d. 4 juli 2005 en van de MBN Voorzitter d.d. 20 maart 2006.

4 BEOORDELINGSPROCEDURE

De beoordeling van de 26 programma-evaluaties die in het vorige hoofdstuk werden geïdentificeerd, vond plaats in een aantal stappen. Dit hoofdstuk beschrijft de werkwijze die daarbij werd gevolgd.

4.1 Beoordelingslijst

Zoals eerder vermeld, is de invulling van het IOB-mandaat met betrekking tot het MFP-breed geënt op de Regeling Prestatiegegevens en Evaluatie-onderzoek. Deze uit 2001 daterende regeling, die op haar beurt een uitwerking vormt van de Comptabiliteitswet, beoogt onder meer vast te leggen aan welke kwaliteitseisen evaluatie-instrumenten behoren te voldoen als zij worden gehanteerd ter verantwoording van de besteding van middelen uit de rijksbegroting. De kwaliteitseisen zijn niet speciaal voor of in het kader van de RPE bedacht: ze werden ontleend aan algemeen gangbare methoden en technieken van sociaal-wetenschappelijk onderzoek.

Uitgangspunt van de RPE is dat de kwaliteit van een evaluatie-onderzoek is vast te stellen aan de hand van de drie criteria *validiteit*, *betrouwbaarheid* en *bruikbaarheid*. Voor de toetsing van evaluatiestudies op hun methodologische kwaliteit dienden deze RPE-criteria verder te worden geoperationaliseerd tot een hanteerbaar beoordelingsinstrument. De ontwikkeling hiervan was reeds geëntameerd in het kader van een eerder IOB-onderzoek.¹³ Hetzelfde stramien, neergelegd in een zogenaamde *Beoordelingslijst*, is toegepast in de onderhavige studie, waarbij in de loop van het onderzoek enkele redactionele aanpassingen en verduidelijkingen werden aangebracht. De basisstructuur van de beoordelingslijst is als volgt voor te stellen:

Figuur 4–1 Beoordelingspiramide



De treden van deze piramide verbeelden de verschillende beoordelingsniveaus, waarbij van boven naar beneden elke laag is te beschouwen als een verdere *operationalisering* van de eraan voorafgaande. De complete Beoordelingslijst met beknopte toelichting op elk van de componenten is opgenomen in bijlage 3. Zo zijn de drie RPE-kwaliteitscriteria Validiteit, Betrouwbaarheid en Bruikbaarheid geoperationaliseerd door middel van respectievelijk zes, vijf en drie *indicatoren*. Op hun beurt zijn deze nader geconcretiseerd in 32 componenten. De bedoeling van deze steeds verdergaande detaillering was om de te hanteren indicaties voor evaluatiekwaliteit steeds *SMART-er*¹⁴ te krijgen. Dat wil dus zeggen: steeds concreter en

¹³ IOB (2004): *Onderzoek naar de kwaliteit van in 2002 afgeronde decentrale evaluaties*. Den Haag: IOB Evaluaties nr. 298.

¹⁴ SMART is een Engelstalig acroniem dat staat voor: Specifiek, Meetbaar, *Attainable* [Bereikbaar, haalbaar], Relevant en Tijdgebonden. Het somt de eigenschappen op waaraan een ideale indicator dient te voldoen.

eenduidiger, hetgeen overigens uiteindelijk slechts in beperkte mate is gelukt. Dat het proces van verdergaande concretisering niet eindeloos kan worden voortgezet, 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 was in praktische zin de grens van de mogelijke detaillering bereikt.

4.2 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, de daarbij betrokken actoren en het onderzoeksrapport. Deze kenmerken spelen zelf geen rol bij de kwaliteitsbeoordeling.

Vervolgens vindt de toetsing plaats aan de drie RPE-criteria: Validiteit, Betrouwbaarheid en Bruikbaarheid. Deze criteria zijn geoperationaliseerd door de definiëring van veertien indicatoren, die elk op hun beurt nader worden gespecificeerd in één of meer concrete componenten. De beoordeling op deze (32) componenten vindt plaats op een 4-puntsschaal 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 tientallen beoordelingsaspecten – van het andere zou kunnen onderscheiden. Een langere schaal, zoals de in het onderwijs veel gebruikte 10-puntsschaal, voert makkelijk tot misleidende schijnnaauwkeurigheid. Daarnaast kent een schaal met een even aantal scoremogelijkheden een duidelijk omslagpunt – in dit geval tussen de tweede en derde score – dat de beoordelaar dwingt tot kiezen. Dit voorkomt dat, bijvoorbeeld uit gemakzucht, het "veilige" middenpunt wordt opgezocht, zoals bij een oneven schaal wel mogelijk is. In de praktijk ziet men dan ook de 'neutrale' middenscore op oneven schalen vaak gebruikt worden in plaats van 'weet niet' 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 volgende uitgangspunten of vuistregels gehanteerd:

- **'slecht'** is in 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 evaluatie-onderzoek, 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.

- **'matig'** is als kwalificatie gehanteerd in die gevallen waarin aan een component niet meer dan rituele aandacht is besteed. Dat wil zeggen dat het betreffende aspect wel wordt *genoemd*, 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 onderzoeksaspect zodanig inhoudelijk aandacht is besteed, dat het positieve effect daarvan zichtbaar is in de uitkomsten van het beoordeelde evaluatie-onderzoek. 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 onderzoeksaspect 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 laatste betekende dat rekening diende te worden gehouden met praktische beperkingen die bijvoorbeeld voortvloeiden uit de aard van de geëvalueerde interventies of de situatie in het betreffende land, enz.

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. Dergelijk 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 zelfs dat van de drie criteria, moet met die waarderingen kunnen worden gerekend. Dat betekent in de eerste plaats dat de nu gebruikte ordinale waarderingsschaal omgezet moet worden naar een hoger meetniveau (interval- of ratioschaal). En dat vervolgens het relatieve gewicht van de individuele componenten of indicatoren ten opzichte van elkaar dient te worden vastgesteld.

Met name de tweede bewerking levert hoofdbrekens op. Hoewel alle 32 beoordelingscomponenten relevant kunnen worden geacht voor de kwaliteit van PE-onderzoek, kunnen zij niet allemaal van evenveel belang voor het kwaliteitsoordeel zijn. Het bepalen van de relatieve gewichten en de onderbouwing daarvan bleek echter, mede vanwege de nog zeer beperkte ervaring met kwaliteitsbeoordeling van evaluatie-onderzoek, niet haalbaar. Ook overleg van IOB met de Werkgroep E&B van het MBN leverde geen bruikbare oplossing voor dit lastige vraagstuk op, zodat van verdere pogingen, althans voor de duur van de lopende MFP subsidieperiode, werd afgezien. De gevolgen voor het gebruiksgemak van de vervaardigde beoordelingen, zeker ten behoeve van de leerfunctie zijn overigens beperkt. Immers verbeteringsmaatregelen zijn het beste uitvoerbaar, als de mogelijkheden ertoe op een zo concreet mogelijk niveau zijn aan te geven. Het is bijvoorbeeld eenvoudiger iets aan de steekproefverantwoording (component 1.6.3) te corrigeren dan "de probleemstelling" (indicator 1.2) of, nog moeilijker: "de validiteit" (criterium 1) te verbeteren (zie bijlage 3).

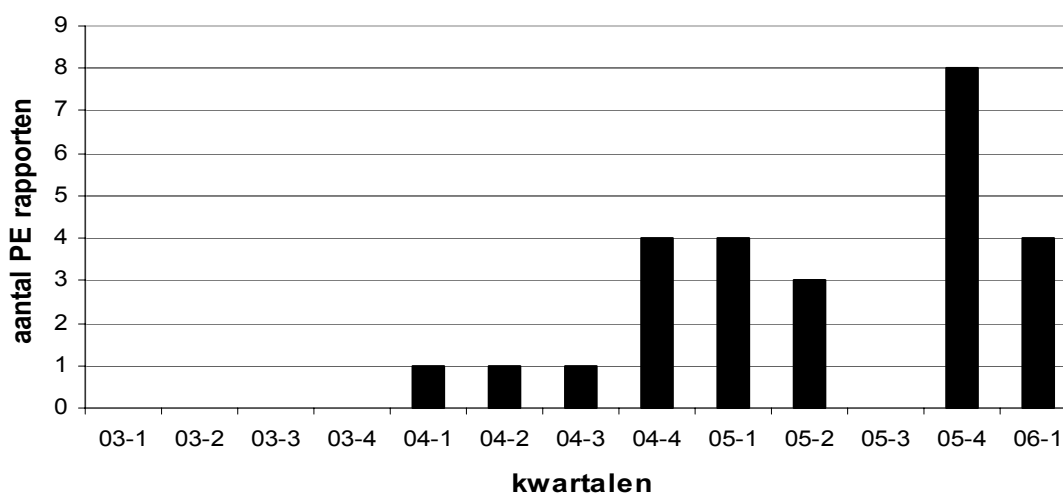
4.3 Beoordelingsprocedure

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

4.3.1 Aanlevering

Het beoordelingstraject begon met de aanlevering van het eindrapport van de betreffende PE bij IOB. Omdat de meerjaren onderzoeksagenda van het MBN pas in 2003 van start ging, en vanwege de aanvankelijke onduidelijkheid over de inhoud van het begrip 'Individuele programma-evaluatie' kwam de aanlevering van rapporten langzaam op gang. Het rapport over de eerste individuele PE werd ontvangen in januari 2004 en dat over de eerste MBN programma-evaluatie in juni van hetzelfde jaar. Ook na die primeurs verliep de toezending niet gelijkmatig, zoals onderstaande grafiek laat zien.

Figuur 4–2 Aanlevering PE-rapporten per kwartaal 2003–2006



Bron: IOB archief MFP PE kwaliteit

Een en ander leverde de, althans voor IOB, ongebruikelijke situatie op dat op het moment dat tot het onderzoek werd besloten, het onderzoeksobject in de vorm van de beoordelen PE-studies nog niet bestond. Evenmin was bekend hoe dat object (zeker voor wat betreft de individuele PE's) eruit zou gaan zien, wat de omvang ervan zou worden, en wanneer het voor onderzoek beschikbaar zou komen.¹⁵

Moest er aanvankelijk vanuit IOB nogal aan de toezending van rapporten worden getrokken, tegen het einde van de beoordelingsperiode versnelde de aanlevering sterk, met als gevolg dat bijna de helft van alle ontvangen rapporten in de laatste twee kwartalen arriveerde. Hierdoor dreigde de discussie met de MFO's over de inhoud van de beoordelingen onder druk te komen.

4.3.2 Concept-beoordeling

Als gevolg van praktische beperkingen was het niet mogelijk om vanuit de Inspectie meer dan één evaluator voor het beoordelingsproces beschikbaar te stellen. Daarom moest op een andere manier worden voorzien in waarborgen tegen het risico van te eenzijdige oordeelsvorming door

¹⁵ Dit had onder meer tot gevolg dat er, voorafgaand aan het onderzoek, geen *Terms of Reference* konden worden geformuleerd.

de IOB-onderzoeker. Hiertoe werd een beoordelingsprocedure in twee rondes (een schriftelijke en een mondelinge) gevolgd. De bedoeling daarvan was de MFO's ruime mogelijkheden te bieden voor commentaar en correcties, vanuit de gedachte dat zij niet alleen zelf het grootste belang hebben bij een zorgvuldige beoordeling van hun werk, maar dat zij ook over de beste kennis beschikken om onzorgvuldigheden te verhinderen.

Na ontvangst van de rapporten bij IOB werden ze eerst bestudeerd en gecontroleerd op volledigheid van de informatie die nodig was om een concept-beoordeling op te kunnen stellen. Ontbrekende informatie werd vervolgens bij de betreffende MFO of, in het geval van MBN-evaluaties, bij de coördinator opgevraagd. Dit bleek ook in elk van de 26 bestudeerde gevallen noodzakelijk. Het leidde dikwijls tot vertraging, omdat de beantwoording van de vragen om aanvullende gegevens vaak nogal wat tijd in beslag nam. Aanvankelijk werden er aan de gestelde vragen door IOB ook observaties toegevoegd over de meest in het oog springende sterke en zwakke kanten, maar dit werd later, omdat de reactietijden zo lang bleken, achterwege gelaten. Die tijden moesten aanvankelijk worden uitgedrukt in maanden (het traagheidsrecord staat op ruim 9 maanden), maar liepen in de loop van 2005 terug van enkele maanden tot een aantal weken. Naarmate de datum voor afsluiting van de gegevensverzameling naderde, werden de reactietijden nog korter (het snelheidsrecord staat op precies één dag).

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 componentscore via een paginaverwijzing aangegeven op welke passage in het rapport of de latere toelichting de betreffende waardering primair, maar niet per se uitsluitend, was gebaseerd. De concept-Beoordeling werd vervolgens aan de indiener van het betreffende rapport toegezonden, vergezeld van het aanbod de gegeven scores mondeling toe te lichten en eventuele verschillen van inzicht omtrent specifieke waarderingen te bespreken.

4.3.3 Bespreking en vaststelling

In vrijwel alle gevallen gingen de MFO's op dit aanbod in. Eigenlijk bleven besprekingen van concept-beoordelingen alleen achterwege als zij door praktische omstandigheden werden verhinderd. Op deze manier vonden ruim 20, doorgaans uitvoerige discussies plaats, waaraan meestal niet alleen evaluatievertegenwoordigers deelnamen, maar ook programmabeheerders die de inhoudelijke verantwoordelijkheid droegen voor de geëvalueerde beleidsonderdelen. Dit bevorderde niet alleen 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.

Behalve voor toelichting op de concept-scores waren de besprekingen, gegeven het legitieme belang van de betrokken MFO-vertegenwoordigers bij de uitkomst van de IOB-beoordelingen, ook expliciet bedoeld om betrokken MFO-vertegenwoordigers de gelegenheid te bieden tegengas te geven. Van de mogelijkheden daartoe is – terecht – vrij intensief gebruik gemaakt om een nadere mondelinge onderbouwing te verlangen van iedere waardering waarvoor zij de overwegingen onduidelijk of onjuist vonden. Hierdoor werd de Inspectie niet alleen gedwongen om te onderbouwen waarom scores, waarvoor men de overwegingen onduidelijk of onterecht vond, precies waren toegekend, maar bijvoorbeeld ook om aan te geven wat men in gegeven omstandigheden had kunnen doen om een hogere waardering te rechtvaardigen. In die gevallen waarin de onderbouwing voor een concept-score niet houdbaar bleek tegenover de argumentatie van MFO of MBN-zijde, werd de waardering herzien. Dit kwam geregeld voor, waarmee het nut van de mondelinge discussieronde werd onderstreept.

Na afronding van deze discussie en aanbrenge van eventueel noodzakelijke wijzigingen werden ten slotte de beoordelingen in definitieve versie vastgesteld en aan de betrokken MFO's toegezonden.

Tot slot

Alles bijeengenomen was de hele beoordelingsprocedure enigszins omslachtig en, waarschijnlijk 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. Blijkens de in het algemeen constructieve opstelling van zowel evaluatie- als programmavertegenwoordigers was de intentie hiertoe bij alle vijf MFO's onmiskenbaar aanwezig. Mede als gevolg van de wat trage start is de periode waarover de PE-beoordelingen zich feitelijk hebben uitgestrekt echter te kort geweest om vast te stellen of leereffecten op de kwaliteit van programma-evaluaties zich al daadwerkelijk hebben voorgedaan.

5 ONDERZOEKRESULTATEN

In dit laatste hoofdstuk wordt weergegeven tot welke uitkomsten de toepassing van het in hoofdstuk 4 beschreven beoordelingssysteem op de 26 onderzochte programma-evaluaties heeft geleid.

De presentatie en analyse van bevindingen wordt *niet* gedifferentieerd naar MFO, om twee redenen. Ten eerste zijn er, zeker na aftrek van de gezamenlijke studies (bij 7 van de 26 PE's was meer dan één MFO betrokken) per individuele organisatie te weinig data voor zinvolle onderlinge vergelijkingen. En ten tweede blijken, voor zover uit het beschikbare materiaal valt op te maken, de overeenkomsten tussen de onderzoeken van de verschillende individuele MFO's aanzienlijk groter dan de verschillen.

In bijlage 4 is een tabel opgenomen die een (geanonimiseerd) overzicht geeft van de scores van de 26 bestudeerde programma-evaluaties op alle 32 beoordelingscomponenten. Behalve dat uit dit overzicht blijkt dat alle PE's hun sterke en zwakke kanten hebben, is het niet makkelijk er duidelijke patronen in te herkennen. Meer inzicht biedt een aggregatie van de beoordelingsuitkomsten per criterium, indicator en component. Daarom zijn de paragrafen 5.3 t/m 5.4 langs die lijnen gestructureerd. Ze worden voorafgegaan door een korte blik op de voornaamste algemene kenmerken van de bestudeerde onderzoeken in § 5.1.

Indien de resultaten met betrekking tot de gezamenlijke MBN-evaluaties als groep duidelijk verschillen van die voor de individuele PE's, wordt dit vermeld. Daarbij dient de lezer dan wel voor ogen te houden dat ook op dit geaggregeerde niveau, gezien het geringe aantal van vier MBN PE's, ondubbelzinnige conclusies niet mogelijk zijn.

5.1 Algemene kenmerken

Bij 16 van de 22 individuele PE's vormden "programma's" het voorwerp van evaluatie (zie § 3.2.1). Hierbij valt in het algemeen te denken aan reeksen activiteiten onder gemeenschappelijk beheer en gefinancierd uit dezelfde bron (binnen het MFO-budget), bijvoorbeeld een landenprogramma. De gezamenlijke programma-evaluaties kenden daarentegen meestal een bredere opzet met, in drie van de vier gevallen, een dwarsdoorsnijdend thema als object van onderzoek. Een onbedoelde, maar niettemin vrij lastige, complicatie van een dergelijke keuze was dat het doorgaans moeilijker bleek de financiële omvang van het evaluatie-object helder te krijgen. Immers, de bedragen voor verrichte uitgaven moeten bij een thematisch opgezette evaluatie uit een groot aantal verschillende budgetten bij elkaar worden 'gesprokkeld'.

Alle MBN PE's bestreken meerdere landen (3 tot 5), terwijl bij de 22 individuele evaluaties de geografische invalshoeken gelijkelijk waren verdeeld over de onderzoeken gericht op één land (8), een regio of groep van landen (7) en wereldwijd of zonder geografische afbakening (7).

Financieel belang

Het financiële belang van PE-objecten, in termen van beslag op MFP-middelen, was doorgaans niet meteen duidelijk. Voor deze studie ging het hierbij, zoals in contacten met de MFO's bij herhaling werd benadrukt, om meer dan een administratief detail, en wel vanuit twee gezichtshoeken: ten eerste bepaalt het financiële belang van het evaluatie-object het uitgavenvolume waarvoor de evaluatieve uitspraken van de beoordeelde onderzoeken gelden; en ten tweede is het gegeven van belang voor de vraag welk deel van de totale MFP-uitgaven door

programma-evaluaties worden bestreken (c.q. afgedekt). In nagenoeg alle gevallen moest na ontvangst van de evaluatierapporten door IOB apart worden geïnformeerd naar de omvang van de geëvalueerde uitgaven, waarbij de beantwoording soms nog wel wat voeten in de aarde had. In een enkel geval bleek de omvang niet met zekerheid vast te stellen en moest worden volstaan met een beredeneerde schatting.

In totaal bestreken de beoordeelde programma-evaluaties voor € 364 mln aan MFP-uitgaven, met een gemiddelde omvang van € 14 mln. Deze bedragen worden echter sterk beïnvloed door één uitschieter van € 145 miljoen (*Evaluation of Novib's Food Security Policy and Practices*, zie bijlage 2). Laat men deze weg dan resteert een totaal van € 219 mln. De gemiddelde financiële omvang van de PE-objecten daalt dan tot € 8,8 mln, variërend van € 168.183 tot € 47.133.000. De vier MBN evaluaties bevonden zich, niet verrassend, in de bovenste regionen van deze bandbreedte, met een minimum van € 13 mln, een maximum van € 47 mln en een gemiddelde van € 24 mln.

Evaluatiedekking

Met deze term wordt bedoeld op het deel van de uitgegeven middelen dat wordt bestreken door of, zoals de RPE het uitdrukt, 'afgedekt' met evaluatie-onderzoek. De tijdvakken die door de 26 beoordeelde evaluaties worden bestreken, vallen in de periode tussen 1993 en 2005 en hebben een gemiddelde lengte van ruim 4 jaar (variërend van twee tot elf). Door deze grote variatie is het niet mogelijk de financiële reikwijdte van de uitgevoerde onderzoeken te relateren aan de MFP-uitgaven in dezelfde begrotingsjaren. Enig perspectief valt slechts aan te brengen door de omvang van de geëvalueerde uitgaven te vergelijken met de hoeveelheid MFP-fondsen die ter beschikking kwamen voor nieuwe uitgaven in dezelfde periode als waarin de evaluatieresultaten werden geproduceerd (2003 – 2006). Als men dit bedrag op circa € 1,6 mrd stelt, bedraagt de omvang van de geëvalueerde uitgaven, afhankelijk van het wel of niet meerekenen van de al genoemde uitschieter, bijna 23 respectievelijk 14% van de nieuw beschikbare middelen.

Onderzoeksuitgaven

Met onderzoeksuitgaven worden hier alleen bedoeld de *out of pocket expenses* ten behoeve van evaluatie-onderzoek, los van en additioneel aan eventuele "vaste" kosten voor evaluatiecapaciteit binnen de MFO, die drukken op de apparaatskostenvergoeding (AKV). Dergelijke vaste kosten hebben voornamelijk betrekking op de loonkosten van eigen evaluatiemedewerkers.

Evenals de geëvalueerde bedragen variëren de uitgaven voor evaluatie-onderzoek sterk, niet alleen in absolute zin, maar ook in verhouding tot het financiële omvang van de geëvalueerde activiteiten. Uitschieters naar boven in absolute bedragen vormen, zoals te verwachten, de MBN-evaluaties die uitgaven van € 174.179 tot 404.973 vergden. De uitgaven ten behoeve van individuele PE liepen uiteen van € 0, in die gevallen waarin een onderzoek helemaal als bureaustudie en door eigen personeel werd uitgevoerd, tot € 110.117 voor een regionale sectorstudie in zes landen.

Om een indruk te geven van de relatieve hoogte van de onderzoeksuitgaven zijn deze in de bijlagen 1 en 2 uitgedrukt in een percentage van het financiële belang van de evaluatie-objecten. Deze percentages fluctueerden, evenals de evaluatie-uitgaven zelf, sterk: van 0,05 tot 8,1. Uitschieters vormden de evaluaties in Oxfam-verband van door Novib mede-gefinancierde programma's. In die gevallen betaalde Novib niet alleen mee aan de uitvoering van het Oxfam programma, maar ook aan de evaluatie ervan. Wanneer de tweede bijdrage proportioneel hoger was dan de eerste, waren de relatieve evaluatie-uitgaven (d.w.z. ten opzichte van de eigen

programmabijdrage) extra hoog. Dit verklaart de percentages van 3,6, 4,0 en 8,1 in bijlage 2. Als men deze uitschieters buiten beschouwing laat variëren de percentages bij de individuele PE's van 0,05 tot 1,5 met een gemiddelde van 0,7.

Er bestaan geen algemeen gehanteerde normen voor de verhouding tussen evaluatie- en programmakosten. Eén uniform percentage is ook moeilijk voor te stellen als het financiële belang van de evaluatie-objecten, zoals in dit onderzoek, uiteenloopt van nog geen 200.000 tot 145 miljoen euro. Tijdens de *Expert Meeting MBN Programma-evaluaties* opperde de voorzitter van de ERG dat “evaluatie[uitgaven] best één procent van je totale [programma]kosten mogen bedragen”. Aan dat streefcijfer voldeden de vier gezamenlijke MBN-evaluaties, waarvoor niet alleen absoluut maar ook relatief meer geld werd uitgetrokken dan voor de meeste individuele PE's, met percentages van respectievelijk 0,9 tot 2,1 (zie bijlage 1). Tweederde van de individuele onderzoeken bleef echter ruimschoots onder deze één procent van de programma-kosten. Vooral wanneer de omvang van dat programma al bescheiden is, biedt een evaluatie-budget van minder dan 1% ervan wel erg weinig ruimte voor het opzetten van gedegen onderzoek. Bijlage 2 bevat hiervan verscheidene voorbeelden. Een en ander roept de vraag op of men niet veeleer een soort glijdende schaal zou moeten hanteren, waarbij de evaluatie-begroting als percentage van de programmakosten daalt naarmate de programma-omvang toeneemt, en omgekeerd. In die gevallen zou een budget voor serieus evaluatie-onderzoek bij een bescheiden programma-omvang significant hoger liggen dan 1% daarvan. Voor zover uit de contacten met de verschillende MFO's duidelijk werd, bieden de huidige evaluatiebudgetten daarvoor geen ruimte.

Onderzoeksdoelstellingen

Hoewel programma-evaluaties meer dan andere vormen van monitoring en evaluatie (M&E) gericht heten te zijn op het afleggen van verantwoording, blijkt uit de bestudering van de PE's niet dat deze functie in de opzet van en rapportage over de betreffende onderzoeken een belangrijke rol heeft gespeeld. Bij de meeste (21 van de 26) evaluaties ligt zowel in de ToR als in het eindrapport de nadruk exclusief op de leerfunctie (in bijlagen 1 en 2 aangeduid met de bredere term *beleidsontwikkeling*).

Slechts in twee van de vier gezamenlijke PE's wordt vermeld dat het onderzoek naast de leerfunctie tevens een verantwoordingsfunctie beoogt te vervullen. Bij de individuele PE's is dat éénmaal het geval, terwijl één andere uitsluitend op de verantwoordingsfunctie is gericht.

Doorlooptijden

Evalueren kost tijd, ook als zich geen onvoorziene obstakels voordoen. Bij de individuele PE's bedroeg de gemiddelde doorlooptijd 7,5 maanden, terwijl de MBN-onderzoeken met gemiddeld 16,5 maanden ruim tweemaal zo lang duurden. In deze laatste categorie werden ten opzichte van de planning in de MBN-onderzoeksagenda aanzienlijke vertragingen opgelopen. Toch kan, wanneer men de omvang van de betreffende exercities in aanmerking neemt, niet worden gesteld dat de gezamenlijk PE's uitzonderlijk veel tijd in beslag hebben genomen. Een juistere constatering lijkt dat de onderzoeksagenda nogal optimistisch was ingevuld.

Sturing en begeleiding

Bij éénderde van de individuele PE's bestond geen structuur of mechanisme voor supervisie of ondersteuning van het onderzoek. De betreffende evaluator(en) was(ren) dus op zichzelf aangewezen. In de overige gevallen werd doorgaans een ad hoc gezelschap geformeerd van programma-verantwoordlijken en evaluatievertegenwoordigers uit de betrokken organisatie(s). Daarmee bleef de onderzoeksbegeleiding dus *intern* van karakter. In geen van de onderzochte gevallen was voorzien in extern toezicht op de methodologische kwaliteit van het evaluatieproces. Hoewel (de beperkte omvang van) het beschikbare materiaal geen harde conclusies over de invloed van de processturing op de kwaliteit van de onderzoeksresultaten toelaat, lijkt het niet overdreven te stellen dat de huidige praktijk op het gebied van sturing en begeleiding ruimte laat voor verbetering.

In tegenstelling tot de individuele PE's genoten alle gezamenlijke PE's onder MBN-vlag wél *externe* begeleiding. Zij ontvingen namelijk, naast een interne begeleidingsstructuur in de voor elke PE speciaal geformeerde Coördinatiegroep (zie § 3.2.2), ondersteuning in de vorm van voornamelijk onderzoeks-methodologisch commentaar en advies in de cruciale stadia van het evaluatieproces. Van buitenaf is niet te beoordelen welke invloed deze begeleiding precies heeft gehad op de uiteindelijke kwaliteit van de vier MBN-evaluaties, maar dat de commentaren op onderzoeksopzet en –rapportage inhoudelijk relevant waren, is aan de hand van de beschikbare teksten makkelijk vast te stellen.

5.2 Validiteit

Bij validiteit draait het om de 'geldigheid' van de onderzoekopzet, en dus om de vraag of de evaluator meet wat hij denkt te meten. Over de validiteit van de bestudeerde programma-evaluaties is een oordeel gevormd aan de hand van een zestal indicatoren, die op hun beurt verder zijn geoperationaliseerd in zeventien 'componenten'. De zes tabellen in deze paragraaf geven de frequentieverdelingen van de waarderingen die aan de afzonderlijke componenten werden toegekend. Deze worden vervolgens becommentarieerd.

Tabel 5–1 Aanleiding en doelstelling

#	component	score			
		g	r	m	s
1.1.1	Vermelding van de aanleiding	9	12	5	0
1.1.2	Duidelijkheid van de (onderzoeks)gebruiksdoelstelling	5	20	1	0

Bron: IOB database MFP PE kwaliteit

Met de formulering van de aanleiding voor evaluatieonderzoek en het doel of de doelen waarvoor de uitkomsten zullen worden gebruikt, doen zich weinig serieuze problemen voor. In de meeste gevallen is de verwoording ruim voldoende tot goed.

Ook de duidelijkheid van de onderzoeksdoelstellingen laat in het algemeen weinig te wensen over. Een veel voorkomende misser is wel het door elkaar halen van de probleemstelling (*wat* men met een onderzoek te weten probeert te komen) en de onderzoeksdoelstelling (*waarvoor* men die wetenschap vervolgens wil gebruiken).

Tabel 5–2 Probleemstelling en onderzoeksvragen

#	component	score			
		g	r	m	s
1.2.1	Helderheid van de probleemstelling	4	18	4	0
1.2.2	Uitwerking van de probleemstelling in onderzoeksvragen	3	11	11	1
1.2.3	Definiëring van evaluatiecriteria	2	6	7	11

Bron: IOB database MFP PE kwaliteit

Waar de formulering van de probleemstelling voor het evaluatie-onderzoek in veruit de meeste gevallen nauwelijks op moeilijkheden stuit, heeft de uitwerking ervan in bruikbare onderzoeksvragen meer voeten in de aarde. De MBN-onderzoeken, waarvan driekwart op dit onderdeel redelijk tot goed scoort, hebben hier overigens minder last van. Een van de meest voorkomende problemen bij de formulering van de onderzoeksvragen – wat opdrachtgevers in veel gevallen aan de geëngageerde evaluatoren zelf overlaten – is een gebrek aan zelfbeperking bij de auteur(s). Zeker wanneer dit wordt gecombineerd met een tekortschietend vermogen om onderscheid te maken tussen hoofd- en bijzaken, kan het resulteren in reeksen (van soms letterlijk tientallen) onderzoeksvragen, waarvan alleen al de lengte het onwaarschijnlijk maakt dat meer dan een klein deel ervan zal worden beantwoord.

In de meeste programma-evaluaties is een ondubbelzinnige definiëring van de gehanteerde evaluatiecriteria geen vanzelfsprekendheid. Integendeel, slechts in uitzonderingsgevallen leggen opdrachtgevers en/of onderzoekers goed en duidelijk vast wat zij onder zulke gangbare begrippen als bijvoorbeeld ‘doeltreffendheid’ of ‘relevantie’ in het kader van hun evaluatie zullen verstaan. Dit is op zichzelf genomen niet zo verwonderlijk, aangezien de termen waarom het gaat, dermate gangbaar zijn in het dagelijkse spraakgebruik, dat de gedachte dat zij mogelijk niet voor alle betrokkenen dezelfde lading hebben, veelal niet eens opkomt. Dit kan fatale gevolgen hebben voor het wederzijdse begrip van de onderzoeksbevindingen, en dus in feite voor de bruikbaarheid ervan.

Tabel 5–3 Evaluatie-object en context

#	component	score			
		g	r	m	s
1.3.1	Omschrijving en afbakening van het onderzoeksobject	4	15	7	0
1.3.2	Beschrijving van de opzet en werking van het onderzoeksobject	2	11	11	2
1.3.3	Situering van het object in de beleidsmatige en institutionele context	5	12	6	3

Bron: IOB database MFP PE kwaliteit

Over het algemeen is de kwalitatieve identificatie van wat de PE beoogt te onderzoeken bevredigend. Dat wil zeggen dat de lezer een behoorlijke voorstelling krijgt van de interventies die geëvalueerd gaan worden. Zoals ook aangegeven in § 5.1. onder ‘Financieel belang’ ontbreekt in veel gevallen wel eenduidige informatie over de hoeveelheid geld die met het evaluatie-object gemoeid is (geweest). Voor de beoordelingen in het kader van dit onderzoek is dat relevant, aangezien die met name plaatsvinden vanuit het verantwoordingsperspectief, en verantwoording afleggen zonder dat vaststaat *waarover* dat precies gebeurt, lastig voorstelbaar is.

Duidelijke uiteenzettingen over de wijze waarop te evalueren activiteiten zijn opgezet en geacht werden te functioneren, ontbreken vaak of laten het nodige te wensen over. Dit geldt met name voor de gezamenlijk PE’s, hetgeen wellicht (mede) is te verklaren uit het feit dat dergelijke onderzoeken bijna altijd veelomvattend zijn en vanuit hun thematische invalshoek een nogal

bonte verzameling van interventies onderzoeken. Hierdoor wordt dit onderdeel van het evaluatieproces bijzonder arbeidsintensief.

De kwaliteit van de contextbehandeling komt er daarentegen over het geheel iets beter af. Opmerkelijk is dat dit mede te danken is aan dezelfde gezamenlijke PE's die onder de vorige component het 'gemiddelde' enigszins omlaag trokken. Bestudering van de onderzoeksrapporten laat ook zien dat in alle MBN-evaluaties veel tot zeer veel nuttige aandacht aan contextfactoren wordt besteed.

Tabel 5-4 Interventiologica

#	component	score			
		g	r	m	s
1.4.1	Weergave of reconstructie van de interventiologica	1	2	13	10
1.4.2	Onderscheiding van resultaatniveaus	3	8	5	10
1.4.3	Operationalisering van resultaatmeting d.m.v. indicatoren	0	2	12	12

Bron: IOB database MFP PE kwaliteit

De beschrijving of reconstructie van de interventietheorie die aan de geëvalueerde activiteiten ten grondslag heeft gelegen, vormt, samen met de operationalisering van de resultaatmeting, het zwakste punt van de meeste programma-evaluaties.

Dit hangt samen met de beoordeling op component 1.3.2 (Opzet en werking van het onderzoeksobject): Zelden treft men in de bestudeerde onderzoeken een duidelijke beschrijving aan van wat nu eigenlijk de bedoeling van de geëvalueerde interventies is geweest, voordat wordt overgegaan tot de weergave van wat ervan terecht is gekomen. De waardering voor de indicatoren reflecteert de mate waarin deze worden toegepast en, zo ja, in hoeverre zij specifiek, meetbaar, bereikbaar [*attainable*], relevant en tijdgebonden (*SMART*) waren. Zoals uit Tabel 5-4 valt af te leiden, was dat eigenlijk nauwelijks het geval.

Aan het onderscheid tussen verschillende resultaatniveaus wordt iets meer aandacht besteed, althans in die zin dat een aantal PE's er blijk van geeft dat opdrachtgevers en/of onderzoekers zich bewust zijn van het verschil tussen *output* en *outcome*. Dit besef vindt echter lang niet altijd een vertaling in een systematische en gedifferentieerde vaststelling van resultaten.

Tabel 5-5 Onderzoeksmethoden

#	component	score			
		g	r	m	s
1.5.1	Specificatie van onderzoeksmethoden	5	20	1	0
1.5.2	Verantwoording van onderzoeksmethoden	2	11	11	2
1.5.3	Vermelding van beperkingen van het onderzoek	5	11	9	1

Bron: IOB database MFP PE kwaliteit

De identificatie c.q. beschrijving van de toegepaste onderzoeksmethoden is over vrijwel de hele linie ruim voldoende of beter. Minder frequent wordt aangegeven *waarom* bepaalde methoden en technieken zijn gehanteerd, wat de eventuele voor- en nadelen ervan waren, en waarom bepaalde voor de hand liggende methoden *niet* zijn toegepast.

Wat beter is het gesteld met de vermelding (en soms ook uitleg) van de beperkingen waaronder het onderzoek is uitgevoerd, voor zover die gevolgen kunnen hebben voor het gebruik van de resultaten. Zeker in de gevallen waarin dit “goed” gebeurt, komt het de geloofwaardigheid van de betreffende PE zeer ten goede.

Tabel 5–6 Consistentie en soliditeit

#	component	score			
		g	r	m	s
1.6.1	Gegevensanalyse en interpretatie, formulering van bevindingen	0	14	12	0
1.6.2	Onderbouwing van conclusies door bevindingen	1	16	9	0
1.6.3	Overeenstemming tussen aanbevelingen en conclusies	0	17	9	0

Bron: IOB database MFP PE kwaliteit

Hoewel de analyse en interpretatie van verzamelde onderzoeksgegevens in ruim de helft van de gevallen redelijk verloopt, doen zich ook wezenlijke tekortkomingen voor. Deze hebben dan voornamelijk betrekking op consistentie en transparantie, waardoor het voor de buitenstaander onduidelijk blijft hoe en waarom specifieke data worden geselecteerd en gecombineerd. Omdat dit soms samengaat met een zekere vaagheid in de gebruikte formuleringen, kan een impressionistisch beeld ontstaan, dat het trekken van heldere conclusies bemoeilijkt.

De onderbouwing van conclusies en de aansluiting van aanbevelingen op die conclusies zijn bij de meerderheid van de evaluaties ruim voldoende. Dat wil zeggen dat de gevolgtrekkingen daadwerkelijk door de geconstateerde feiten worden ondersteund en dat aanbevelingen logisch uit de getrokken conclusies voortvloeien. Daarnaast komen echter, in een belangrijke minderheid van de onderzochte evaluaties, op deze punten gedachtesprongen voor die niet te volgen zijn, waarbij conclusies en aanbevelingen zijn geformuleerd die niet door het verrichte onderzoek kunnen worden gedragen. De gezamenlijke MBN PE's hebben van deze tekortkomingen overigens minder last dan de individuele onderzoeken.

5.3. Betrouwbaarheid

Bij de beoordeling van de betrouwbaarheid staat de vraag naar de stabiliteit van de onderzoeksresultaten centraal. Simpel gezegd: in hoeverre komt er, als hetzelfde evaluatie-object nogmaals en met gebruikmaking van dezelfde methoden wordt onderzocht, hetzelfde uit? Voor de toetsing van de programma-evaluaties aan dit criterium zijn vijf indicatoren gehanteerd, die vervolgens nader zijn geconcretiseerd in negen ‘componenten’.

Tabel 5–7 Gegevensbronnen

#	component	score			
		g	r	m	s
2.1.1	Identificatie van gebruikte gegevensbronnen	4	17	5	0
2.1.2	Verificatie van betrouwbaarheid/nauwkeurigheid van gegevensbronnen	1	9	5	11

Bron: IOB database MFP PE kwaliteit

Gegevensbronnen zijn over het algemeen behoorlijk vermeld. Opsomming van materiaal dat door de evaluatoren is gebruikt of althans aan hen ter beschikking heeft gestaan, is vaak uitvoerig. Wel komt het slechts zelden voor dat bevindingen in de onderzoeksrapportage, bijvoorbeeld in voetnoten, met specifieke schriftelijke of menselijke bronnen in verband worden gebracht.

Expliciete controle van bronnen, bijvoorbeeld door middel van triangulatie, komt maar in beperkte mate voor. In de onderzoeksrapportages is weinig plaats voor scepsis en verificatie. Dit is relevant omdat het overgrote deel van de geraadpleegde documentatie en informanten afkomstig is van c.q. behoort tot partijen die een reëel – en overigens legitiem – belang hebben bij de uitkomsten van het evaluatie-onderzoek.

Tabel 5–8 Representativiteit

#	component	score			
		g	r	m	s
2.2.1	Verantwoording van onderzochte steekproef c.q. case study selectie	4	6	9	7

Bron: IOB database MFP PE kwaliteit

Bij programma-evaluaties kan doorgaans maar een klein deel van de ‘gevallen’ die tot het onderzoeksobject behoren daadwerkelijk worden onderzocht. Omdat de uitkomsten van PE’s bestemd zijn om verantwoording af te leggen of lessen te trekken op beleidsniveau, is het van belang dat de bevindingen van de onderzochte gevallen ook gelden voor de niet bestudeerde cases (ook wel aangeduid als ‘externe validiteit’). Bij de meeste onderzochte evaluaties doen zich op dit punt meer of minder ernstige tekortkomingen voor. Het beeld dat wordt weergegeven in Tabel 5–8 is in zoverre geflatteerd dat bij drie van de vier PE’s de waardering “goed” in feite berust op de omstandigheid dat hun onderzoekspopulaties zo klein waren, dat die in hun geheel konden worden onderzocht en er dus helemaal geen steekproef hoefde te worden getrokken.

In de overige 23 onderzoeken beperkte de onderbouwing van de case study selectie, als die niet geheel ontbrak, zich tot factoren als ‘gegevensdichtheid’, ‘diversiteit’ en dergelijke. Statistische representativiteit werd echter nergens als doel (of probleem) bij het onderzoeksontwerp aan de orde gesteld. Het gevolg is dat het niet duidelijk wordt hoe generaliseerbaar de onderzoeksuitkomsten zijn, zonder dat de opdrachtgevers dit kennelijk als een obstakel voor de bruikbaarheid ervaren. Wellicht mede als gevolg van hun relatief grote omvang hebben de gezamenlijke PE’s nog iets meer last van dit probleem dan de individuele.

Tabel 5–9 Onafhankelijkheid

#	component	score			
		g	r	m	s
2.3.1.	Onafhankelijkheid van gegevensbronnen	0	10	14	2
2.3.2	Onafhankelijkheid van onderzoekers	0	23	3	0

Bron: IOB database MFP PE kwaliteit

Hoewel de hoeveelheid referentiemateriaal vaak indrukwekkend is en de variatie binnen de kring van belanghebbende partijen redelijk gevarieerd, komt het gebruik van volledig onafhankelijk materiaal nauwelijks voor.

Voor de ingeschakelde onderzoekers geldt dat die nooit *volledig* onafhankelijk *kunnen* zijn, omdat zij uiteindelijk door de betrokken MFO’s als opdrachtgevers moeten worden geëngageerd en gedechargeerd (en betaald). Met inachtneming van die onvermijdelijke beperking kan worden vastgesteld dat de MFO’s zich serieus hebben ingespannen om gekwalificeerde, externe deskundigen aan te trekken en daarin ook vaak zijn geslaagd. Hierbij is wel op te merken dat bij de rekrutering van evaluatoren vaak meer aandacht wordt besteed aan hun materie–

deskundigheid dan aan zuivere (sociaal-wetenschappelijke) onderzoeksexpertise, hetgeen soms ten koste kan gaan van de vereiste distantie van de onderzoeker tot het onderzoeksobject.

Tabel 5–10 Onderzoeksverloop

#	component	score			
		g	r	m	s
2.4.1	Uitvoering binnen geplande tijd en beschikbaar budget	8	12	6	0
2.4.2	Verantwoording van het onderzoeksverloop	6	13	6	1

Bron: IOB database MFP PE kwaliteit

Over het algemeen werd heel redelijk de hand gehouden aan de overeengekomen tijdplanning en bleven de uitgaven binnen de beschikbaar gestelde onderzoeksbegroting. Met name met de tijdplanning hadden de MBN-onderzoeken, ongetwijfeld als gevolg van hun omvang en complexiteit, ‘gemiddeld’ wat meer moeite dan de individuele programma-evaluaties.

Tabel 5–11 Kwaliteitsbewaking

#	component	score			
		g	r	m	s
2.5.1	Verwerking van commentaar van belanghebbenden	5	16	4	1
2.5.2	Intern of extern toezicht door begeleidings- of stuurgroep	3	12	7	4

Bron: IOB database MFP PE kwaliteit

In veruit de meeste gevallen werd aan belanghebbenden ruim voldoende gelegenheid geboden om commentaar te leveren op concept-bevindingen en –conclusies, maar hoe de verwerking van dat commentaar is gerealiseerd werd meestal niet precies duidelijk. Slechts in een enkel geval voegden de onderzoekers ontvangen commentaar met hun gedetailleerde reactie erop integraal bij hun rapportage.

Als individuele PE’s begeleid of gesuperviseerd werden, anders dan alleen door de directe opdrachtgever, gebeurde dat meestal door een klein, intern gezelschap van programma-verantwoordelijken en/of evaluatievertegenwoordigers, dat betrekkelijk informeel opereerde. Uit de bestudering van de PE-onderzoeken bleek dat in nogal wat gevallen met name methodologische problemen via dergelijke vormen van begeleiding niet werden aangepakt, of zelfs niet opgemerkt. Dat lag volkomen anders bij de gezamenlijke MBN PE’s die niet alleen onder toezicht stonden van een de eigen Coördinatiegroep, maar ook nog te maken kregen met de Evaluatie Referentie Groep, die niet alleen deskundig bleek, maar daarnaast geen blad voor de mond nam als het erom ging opdrachtgevers en evaluatoren voor dwalingen te behoeden. Voor zover viel na te gaan, heeft de ERG hiermee een heilzame invloed op de kwaliteit van de MBN-evaluaties uitgeoefend.

5.4 Bruikbaarheid

Ten slotte wordt een oordeel gevormd over de bruikbaarheid van de onderzochte PE’s. Die wordt uiteraard mede bepaald door de hiervoor behandelde validiteit en betrouwbaarheid, maar kent ook enkele eigen ‘meetpunten’. Deze zijn geconcretiseerd in een drietal indicatoren, die vervolgens zijn gesplitst in zes ‘componenten’. Voor de goede orde zij er nog op gewezen dat *bruikbaarheid* in de praktijk geen voorwaarde hoeft te zijn voor daadwerkelijk *gebruik* van de resultaten van een onderzoek. Ook de uitkomsten van kwalitatief-methodologisch ‘onbruikbaar’ onderzoek *kunnen* een ‘gebruiker’ van pas komen.

Tabel 5–12 Aansluiting op behoeften

#	component	score			
		g	r	m	s
3.1.1	Eenduidigheid en toegankelijkheid van onderzoeksresultaten	1	16	9	0
3.1.2	Operationaliteit en realisme van aanbevelingen	2	14	10	0

Bron: IOB database MFP PE kwaliteit

De meeste PE-resultaten waren voldoende duidelijk en begrijpelijk, maar in een significante minderheid van de onderzoeken vertoonden de uitkomsten op deze punten toch tekortkomingen. Deze hingen, zoals te verwachten, vooral samen met of vloeiden voort uit gebreken die onder Tabel 5–6 werden gesignaleerd bij de gegevensanalyse e.d.

Bij de meerderheid van onderzoeken lagen de gedane aanbevelingen binnen het bereik van de betrokken programma-verantwoordelijken. In een aantal gevallen vergde de opvolging van aanbevelingen echter verstreckende maatregelen of faciliteiten die zich niet onder de controle van de verantwoordelijke opdrachtgevers bevonden of die (helemaal) niet door de evaluatoren waren onderzocht. Hierdoor kregen de betreffende aanbevelingen een vrijblijvend karakter.

Tabel 5–13 Volledigheid

#	component	score			
		g	r	m	s
3.2.1	Volledigheid van beantwoording van onderzoeksvragen	1	15	10	0
3.2.2	Verklaring van verschillen tussen beoogde en behaalde resultaten	0	4	3	19

Bron: IOB database MFP PE kwaliteit

Voor het grootste deel van de PE's was de beantwoording van de onderzoeksvragen (meestal door de conclusies van de evaluatie) ruim voldoende, maar er kwamen ook opvallende missers voor. Dat laatste was met name het geval als het onderzoeksontwerp overbelast was geraakt door het onder Tabel 5–2 reeds gesignaleerde probleem van de overdaad aan onderzoeksvragen. In die gevallen bleken onderzoeksvragen overigens vaak te kunnen worden genegeerd zonder dat dat tot kritische reacties van opdrachtgevers leidde.

Tot een steekhoudende verklaring van afwijkingen tussen geplande en gerealiseerde uitkomsten van geëvalueerde interventies kwam het in de bestudeerde onderzoeken vrijwel nooit. Dat was, mede gezien de onder de Tabellen 5–3 en 5-4 gesignaleerde problemen met de weergave van de werking van de evaluatie-objecten en de reconstructie van de interventielogica, waarschijnlijk ook niet te verwachten.

Tabel 5–14 Presentatie

#	component	score			
		g	r	m	s
3.3.1	Helderheid en representativiteit van samenvatting	5	13	3	5
3.3.2	Scheiding en onderscheid tussen aanbevelingen en conclusies	11	10	5	0

Bron: IOB database MFP PE kwaliteit

De bruikbaarheid van samenvattingen varieerde nogal sterk: Bij ongeveer éénvijfde van de onderzoeken waren ze zonder meer “goed”, terwijl ze in evenveel gevallen volledig ontbraken. In de meeste gevallen waren ze echter ruim voldoende.

De fysieke scheiding tussen conclusies en aanbevelingen was in veruit de meeste gevallen duidelijk. Nog belangrijker was wellicht dat het *onderscheid* tussen deze twee traditionele elementen van evaluatie-onderzoek in de eindrapportage helder bleef.

BIJLAGEN

BIJLAGE 1: KENMERKEN MBN PROGRAMMA-EVALUATIES

EVALUATIEOBJECT	
titel	The Role of Women's Organisations in Civil Society Building
mfo's	Cordaid¹⁶ – Hivos – Icco
type	thema
land(en), regio('s)	Kenya, Zimbabwe, Zuid-Afrika
financieel belang (€)	13.403.600
evaluatieperiode	1998 – 2003
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling + verantwoording
type	evaluatie (ex post)
kosten (€)	276.210
kosten/belang (%)	2,1
doorlooptijd (mnd)	18
sturing	Coördinatiegroep + ERG
EVALUATIERAPPORT	
datum	november 2004
omvang (pp)	156

EVALUATIEOBJECT	
titel	Evaluation of Health Theme
mfo's	Cordaid – Icco – Plan
type	sector
land(en), regio('s)	Congo, Nepal, Zambia
financieel belang (€)	13.262.901
evaluatieperiode	2002 – 2004
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	evaluatie (ex post)
kosten (€)	174.179
kosten/belang (%)	1,3
doorlooptijd (mnd)	17
sturing	Coördinatiegroep + ERG
EVALUATIERAPPORT	
datum	oktober 2005
omvang (pp)	194

¹⁶ Bij iedere gezamenlijke programma-evaluaties vervulde één van de deelnemende MFO's de voortrekkerrol. In deze overzichten is die MFO vetgedrukt weergegeven.

EVALUATIEOBJECT	
titel	Assessing Civil Society Participation
mfo's	Cordaid – Hivos – Novib – Plan
type	thema
land(en), regio('s)	Colombia, Guatemala, Sri Lanka, Oeganda
financieel belang (€)	47.133.000
evaluatieperiode	1999 – 2004
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	evaluatie (ex post)
kosten (€)	404.973
kosten/belang (%)	0,9
doorlooptijd (mnd)	17
sturing	Coördinatiegroep + ERG
EVALUATIERAPPORT	
datum	december 2005
omvang (pp)	645

EVALUATIEOBJECT	
titel	MBN HIV/AIDS Evaluation
mfo's	Cordaid, Hivos, Icco, Novib , Plan
type	thema
land(en), regio('s)	India, Malawi, Zimbabwe, Zuid-Afrika
financieel belang (€)	22.687.000
evaluatieperiode	2001 – 2004
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling + verantwoording
type	evaluatie (ex post)
kosten (€)	375.000
kosten/belang (%)	1,7
doorlooptijd (mnd)	14
sturing	Coördinatiegroep + ERG
EVALUATIERAPPORT	
datum	februari 2006
omvang (pp)	646

BIJLAGE 2: KENMERKEN INDIVIDUELE PROGRAMMA-EVALUATIES

EVALUATIEOBJECT	
titel	Evaluation Regional and Support Organisation (RSO) Programme
mfo's	Cordaid – Icco
type	programma
land(en), regio('s)	Midden- en Oost-Europa, Kaukasus en Centraal-Azië
financieel belang (€)	340.000
evaluatieperiode	1999 – 2001
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	evaluatie (ex post)
kosten (€)	geen
kosten/belang (%)	nvt (overhead)
doorlooptijd (mnd)	13
sturing	geen
EVALUATIERAPPORT	
datum	februari 2003
omvang (pp)	33
EVALUATIEOBJECT	
titel	'We are All Walking Together': Evaluative Enquiry of I/C Consult
mfo's	Cordaid – Icco
type	instrument
land(en), regio('s)	wereldwijd
financieel belang (€)	3.484.995
evaluatieperiode	1998 – 2003
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	evaluatie (ex post)
kosten (€)	48.000
kosten/belang (%)	1,3
doorlooptijd (mnd)	10
sturing	interne begeleidingsgroep
EVALUATIERAPPORT	
datum	oktober 2003
omvang (pp)	56

EVALUATIEOBJECT	
titel	Sustainable Agriculture Programme of VECO-Indonesia
mfo	Cordaid
type	programma
land(en), regio('s)	Indonesië
financieel belang (€)	168.183
evaluatieperiode	2001 – 2004
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	evaluatie (interim)
kosten (€)	0
kosten/belang (%)	nvt
doorlooptijd (mnd)	8
sturing	interne begeleidingsgroep + externe (multidonor) stuurgroep
EVALUATIERAPPORT	
datum	juni 2005
omvang (pp)	105

EVALUATIEOBJECT	
titel	Incubation for Change: Evaluation of ASA Bangladesh
mfo	Cordaid
type	programma
land(en), regio('s)	Bangladesh
financieel belang (€)	1.000.000 (ca.)
evaluatieperiode	2001 – 2004
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	review
kosten (€)	7.582
kosten/belang (%)	0,8
doorlooptijd (mnd)	5
sturing	interne begeleidingsgroep
EVALUATIERAPPORT	
datum	februari 2005
omvang (pp)	58

EVALUATIEOBJECT	
titel	Cordaid Programma Suriname
mfo	Cordaid
type	programma
land(en), regio('s)	Suriname
financieel belang (€)	6.600.000
evaluatieperiode	1999 – 2004
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	evaluatie (ex post)
kosten (€)	45.000
kosten/belang (%)	0,7
doorlooptijd (mnd)	14
sturing	interne begeleidingsgroep
EVALUATIERAPPORT	
datum	december 2005
omvang (pp)	150

EVALUATIEOBJECT	
titel	Review of the Hivos Central Asia programme
mfo	Hivos
type	programma
land(en), regio('s)	Kazachstan, Kirgizië
financieel belang (€)	7.100.000
evaluatieperiode	1994 – 2002
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	review
kosten (€)	30.000
kosten/belang (%)	0,4
doorlooptijd (mnd)	3
sturing	interne begeleidingsgroep
EVALUATIERAPPORT	
datum	september 2003
omvang (pp)	46

EVALUATIEOBJECT	
titel	Access for All: Equal Opportunities in Cyberspace
mfo	Hivos
type	programma
land(en), regio('s)	wereldwijd
financieel belang (€)	7.015.442
evaluatieperiode	2000 – 2003
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	evaluatie (ex post)
kosten (€)	45.000
kosten/belang (%)	0,6
doorlooptijd (mnd)	9
sturing	interne begeleidingsgroep
EVALUATIERAPPORT	
datum	augustus 2004
omvang (pp)	66

EVALUATIEOBJECT	
titel	Evaluation of Hivos Film Sector in Asia and South-Eastern Europe
mfo	Hivos
type	programma
land(en), regio('s)	Nederland, Bosnië, India, Sri Lanka, Indonesië, Nepal
financieel belang (€)	2.000.000 (ca.)
evaluatieperiode	1993 – 2004
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	evaluatie (ex post)
kosten (€)	30.000
kosten/belang (%)	1,5
doorlooptijd (mnd)	17
sturing	interne begeleidingsgroep
EVALUATIERAPPORT	
datum	januari 2005
omvang (pp)	112

EVALUATIEOBJECT	
titel	Evaluación de los Efectos Directos del Programa de Apoyo de HIVOS a Contrapartes del Movimiento Ambientalista en América Latina
mfo	Hivos
type	sector
land(en), regio('s)	Latijns-Amerika
financieel belang (€)	7.023.270
evaluatieperiode	1999 – 2004
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	evaluatie (ex post) + formulering
kosten (€)	44.253
kosten/belang (%)	0,6
doorlooptijd (mnd)	5
sturing	interne begeleidingsgroep
EVALUATIERAPPORT	
datum	maart 2005
omvang (pp)	326

EVALUATIEOBJECT	
titel	Agrobiodiversity in Action
mfo's	Hivos – Novib
type	programma
land(en), regio('s)	wereldwijd
financieel belang (€)	6.765.334
evaluatieperiode	2000 – 2002
EVALUATIEONDERZOEK	
doelstelling	besluitvorming over voortzetting
type	evaluatie + formulering
kosten (€)	14.017
kosten/belang (%)	0,2
doorlooptijd (mnd)	1
sturing	interne begeleidingsgroep
EVALUATIERAPPORT	
datum	april 2004
omvang (pp)	90

EVALUATIEOBJECT	
titel	Samen (of) eigenzinnig
mfo	lcco
type	programma
land(en), regio('s)	wereldwijd
financieel belang (€)	1.000.000 (ca.)
evaluatieperiode	2000 – 2002
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	evaluatie (interim)
kosten (€)	1.000
kosten/belang (%)	0,1
doorlooptijd (mnd)	0,3
sturing	geen
EVALUATIERAPPORT	
datum	maart 2003
omvang (pp)	15

EVALUATIEOBJECT	
titel	Organisational Development Supported by ICCO
mfo	lcco
type	thema
land(en), regio('s)	wereldwijd
financieel belang (€)	1.964.933
evaluatieperiode	2003 – 2004
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	evaluatie (interim)
kosten (€)	0
kosten/belang (%)	nvt
doorlooptijd (mnd)	8
sturing	afd. Onderzoek en Evaluatie
EVALUATIERAPPORT	
datum	december 2004
omvang (pp)	43

EVALUATIEOBJECT	
titel	ICCO Country Evaluation Philippines
mfo	lcco
type	programma
land(en), regio('s)	Filippijnen
financieel belang (€)	16.620.614
evaluatieperiode	2001 – 2005
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	evaluatie (ex post)
kosten (€)	38.491
kosten/belang (%)	0,2
doorlooptijd (mnd)	9
sturing	geen
EVALUATIERAPPORT	
datum	januari 2006
omvang (pp)	158

EVALUATIEOBJECT	
titel	Violence Against women in South Africa
mfo	Novib
type	programma
land(en), regio('s)	Zuid-Afrika
financieel belang (€)	3.005.267
evaluatieperiode	1996 – 2003
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	evaluatie (interim)
kosten (€)	37.038
kosten/belang (%)	1,2
doorlooptijd (mnd)	5
sturing	interne begeleidingsgroep
EVALUATIERAPPORT	
datum	juli 2003
omvang (pp)	54

EVALUATIEOBJECT	
titel	Evaluation of Oxfam International's Coffee Campaign
mfo	Novib
type	project (campagne)
land(en), regio('s)	wereldwijd
financieel belang (€)	370.000
evaluatieperiode	2002 – 2003
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	evaluatie (ex post)
kosten (€)	30.000
kosten/belang (%)	8,1
doorlooptijd (mnd)	4
sturing	geen
EVALUATIERAPPORT	
datum	mei 2004
omvang (pp)	27

EVALUATIEOBJECT	
titel	Oxfam Mekong Initiative (OMI)
mfo	Novib
type	programma
land(en), regio('s)	Mekong regio
financieel belang (€)	222.862
evaluatieperiode	1997 – 2003
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	review
kosten (€)	7.922
kosten/belang (%)	3,6
doorlooptijd (mnd)	1
sturing	interne (Oxfam) begeleidingsgroep
EVALUATIERAPPORT	
datum	september 2003
omvang (pp)	59

EVALUATIEOBJECT	
titel	Joint Oxfam Programme Malawi (JOPM)
mfo	Novib
type	programma
land(en), regio('s)	Malawi
financieel belang (€)	2.464.026
evaluatieperiode	2001 – 2004
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	review
kosten (€)	10.933
kosten/belang (%)	0,44
doorlooptijd (mnd)	8
sturing	geen
EVALUATIERAPPORT	
datum	januari 2004
omvang (pp)	78

EVALUATIEOBJECT	
titel	Evaluation of Novib's Food Security Policy and Practices
mfo	Novib
type	thema
land(en), regio('s)	wereldwijd
financieel belang (€)	145.000.000
evaluatieperiode	1997 – 2004
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	evaluatie (interim)
kosten (€)	70.000
kosten/belang (%)	0,05
doorlooptijd (mnd)	12
sturing	interne begeleidingsgroep
EVALUATIERAPPORT	
datum	september 2005
omvang (pp)	274

EVALUATIEOBJECT	
titel	Joint Oxfam HIV and AIDS Programme in South Africa (JOHAP)
mfo	Novib
type	programma
land(en), regio('s)	Zuid-Afrika
financieel belang (€)	375.466
evaluatieperiode	2002 – 2004
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	evaluatie (ex post)
kosten (€)	15.000
kosten/belang (%)	4,0
doorlooptijd (mnd)	13
sturing	intern (Oxfam) management team
EVALUATIERAPPORT	
datum	maart 2005
omvang (pp)	88

EVALUATIEOBJECT	
titel	Plan Bolivia: Initial Country Programme Assessment
mfo	Plan
type	programma
land(en), regio('s)	Bolivia
financieel belang (€)	45.000.000
evaluatieperiode	1999 – 2003
EVALUATIEONDERZOEK	
doelstelling	verantwoording
type	evaluatie (meta)
kosten (€)	44.548
kosten/belang (%)	0,1
doorlooptijd (mnd)	7
sturing	geen
EVALUATIERAPPORT	
datum	maart 2004
omvang (pp)	87

EVALUATIEOBJECT	
titel	ROCCA Quality Learning (QL I)
mfo	Plan
type	programma
land(en), regio('s)	El Salvador, Guatemala, Haïti, Honduras, Nicaragua, Dominicaanse Rep.
financieel belang (€)	6.818.462
evaluatieperiode	2001 – 2003
EVALUATIEONDERZOEK	
doelstelling	beleidsontwikkeling
type	evaluatie (ex post)
kosten (€)	110.117
kosten/belang (%)	1,6
doorlooptijd (mnd)	5
sturing	interne begeleidingsgroep
EVALUATIERAPPORT	
datum	december 2003
omvang (pp)	616

EVALUATIEOBJECT	
titel	Searching for Elements of Impact in the Domain of primary Education
mfo	Plan
type	sector
land(en), regio('s)	Sierra Leone, Guatemala
financieel belang (€)	3.500.000
evaluatieperiode	2000 – 2005
EVALUATIEONDERZOEK	
doelstelling	verantwoording
type	evaluatie (ex post)
kosten (€)	40.000
kosten/belang (%)	1,1
doorlooptijd (mnd)	10
sturing	geen
EVALUATIERAPPORT	
datum	oktober 2005
omvang (pp)	131

BIJLAGE 3: BEOORDELINGSLIJST PROGRAMMA-EVALUATIES

Deze beoordelingslijst is gehanteerd om de opzet en uitvoering van Programma-evaluaties (PE's)¹⁷ binnen het Medefinancieringsprogramma (MFP) 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	
onderzoekskosten in € (en in %)	
financieringsbron	
doorlooptijd	
ACTOREN	
opdrachtgever	
sturing of begeleiding	
evaluatoren	• •
vermelding kwalificaties	
explicitering selectiecriteria	
EVALUATIERAPPORT	
datum	
auteur(s)	
onderzoeksvraagstelling	
ToR	
omvang	

¹⁷ Deze omvatten zowel individuele (of 'interne') PE's van medefinancieringsorganisaties (MFO's) als gezamenlijke PE's ondernomen in het kader van de samenwerking in het Medefinancieringsprogramma-Breed Netwerk (MBN).

BEOORDELINGSCRITERIA, INDICATOREN EN SCORES

1	VALIDITEIT <i>(meet men wat men beoogt te meten en is het meetresultaat generaliseerbaar?)</i>	score
1.1	Aanleiding en doelstelling	
1.1.1	Vermelding van de aanleiding voor de evaluatie	
1.1.2	Duidelijkheid van de (onderzoeks)gebruiksdoelstelling	
1.2	Probleemstelling en onderzoeksvragen	
1.2.1	Helderheid van de probleemstelling	
1.2.2	Uitwerking van de probleemstelling in relevante onderzoeksvragen	
1.2.3	Definiëring van de gehanteerde evaluatiecriteria	
1.3	Evaluatieobject en context	
1.3.1	Definiëring en afbakening van het onderzoeksobject	
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	Interventiologica	
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 onderzoeksmethoden	
1.5.2	Verantwoording van de toegepaste methoden	
1.5.3	Vermelding van de beperkingen van het onderzoek	
1.6	Consistentie en soliditeit	
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 (<i>zijn de onderzoeksuitkomsten stabiel?</i>)	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. case study 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	Onderzoeksverloop	
2.4.1	Uitvoering van de evaluatie binnen de geplande tijd en het beschikbare budget	
2.4.2	Verantwoording van het onderzoeksverloop	
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 (<i>zijn de onderzoeksuitkomsten toepasbaar?</i>)	score
3.1	Aansluiting	
3.1.1	Eenduidigheid en toegankelijkheid van de onderzoeksresultaten	
3.1.2	Operationaliteit en realisme van aanbevelingen of lessen	
3.2	Volledigheid	
3.2.1	Beantwoording van de onderzoeksvragen 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	

TOELICHTING OP INVULLING VELDEN

INVENTARISATIE VAN KENMERKEN

titel	volgens de definitieve versie van het eindrapport
type evaluatieobject	bijv. programma, project, sector, thema, instrument of organisatie
land(en), regio('s)	gebied(en) waar het onderzoek betrekking op heeft
financiële omvang	de hoeveelheid geld die door de betreffende MFO('s) tijdens de onderzochte periode aan het evaluatieobject is besteed
evaluatieperiode	de periode die het onderzoek bestrijkt
gebruiksdoelstelling	beleidsontwikkeling (leren), verantwoording of beide
soort evaluatieonderzoek	bijv. ex post of interim; evaluatie of review; evaluatie & formulering ¹⁸ ; meta
onderzoekskosten	in € (en in % van de financiële omvang 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
vermelding kwalificaties	voornaamste opleidings- en ervaringsgegevens van de onderzoekers
explicitering selectiecriteria	vermelding van maatstaven aan de hand waarvan de onderzoekers zijn geselecteerd (bijvoorbeeld in de ToR)
datum	van de definitieve versie van het eindrapport
auteur(s)	aangeven of dit de evaluatoren zelf zijn geweest of anderen
onderzoeksvraagstelling	worden de onderzoeksvragen expliciet <u>in</u> de hoofdtekst gesteld c.q. daarin (uit de ToR) herhaald?
ToR	volledig of verkort (of helemaal niet)– als bijlage in het rapport opgenomen
omvang	van het rapport in pagina's (inclusief bijlagen)

BEOORDELINGSCRITEIA EN -INDICATOREN

validiteit	
1.1.1	de aanleiding vormt de beweegreden van <u>waaruit</u> de evaluatie wordt ondernomen; deze gaat dus aan de doelstelling <u>vooraf</u>
1.1.2	de gebruiksdoelstelling geeft aan <u>waartoe</u> de evaluatie wordt ondernomen, dus <u>waarvoor</u> de resultaten van het onderzoek zullen worden gebruikt
1.2.1	de probleemstelling definieert <u>wat</u> men met het onderzoek te weten wil komen
1.2.2	bij een resultaatgerichte evaluatie mogen bij de uitwerking van de probleemstelling in ieder geval onderzoeksvragen over doeltreffendheid en doelmatigheid worden verwacht
1.2.3	eenduidige omschrijving van de gebruikte maatstaven (bijvoorbeeld doeltreffendheid), zodat helder is <u>waarop</u> het evaluatieobject precies wordt beoordeeld

¹⁸ Van deze combinatie is sprake als de onderzoeksopdracht zodanige aanwijzingen m.b.t. de opstelling van aanbevelingen voor de voortzetting van de te evalueren activiteiten bevat, dat het risico bestaat dat de formulering van deze aanbevelingen de evaluatieve bevindingen van de onderzoekers gaat sturen in plaats van omgekeerd.

1.3.1	ontplooiende activiteiten in de tijd, geografisch, financieel, type interventies, doelgroepen, enz.
1.3.2	weergave van de manier waarop de uitvoering van de onderzochte activiteiten was opgezet en hoe die uitvoering in de praktijk is verlopen
1.3.3	weergave van relevante beleidsmatige achtergronden en uitgangspunten, alsmede van het institutionele/organisatorische krachtenveld waarin het evaluatieobject opereerde
1.4.1	weergave van de beoogde resultaten van geëvalueerde interventies en van hun doorwerking
1.4.2	bevat de gehanteerde interventielogica een doel-middelenhiërarchie met oorzaak-gevolg-relaties?
1.4.3	mate waarin de voor de diverse resultaatniveaus gedefinieerde indicatoren SMART ¹⁹ zijn
1.5.1	duidelijke identificatie van methoden en technieken voor gegevensverzameling en -verwerking
1.5.2	beargumentering van de gebruikte onderzoeksmethoden en -technieken
1.5.3	uitleg over eventuele tekortkomingen van het onderzoek en (expliciete) aanduiding van datgene waarvoor de uitkomsten <u>niet</u> kunnen worden gebruikt
1.6.1	soliditeit van gegevensverzameling en -verwerking
1.6.2	mate waarin de conclusies <u>rechtstreeks</u> zijn getrokken uit de bevindingen
1.6.3	mate waarin de aanbevelingen <u>daadwerkelijk</u> voortvloeien uit de getrokken conclusies
betrouwbaarheid	
2.1.1	documentatie, respondenten, literatuur, enz.
2.1.2	geeft de rapportage blijk van een kritische houding t.o.v. de gebruikte bronnen, bijvoorbeeld doordat is geprobeerd hun waarde te controleren; zijn er meer bronnen of methoden gebruikt om dezelfde gegevens te verzamelen?
2.2.1	beargumentering van de steekproefrepresentativiteit resp. de case study selectie
2.3.1	eventuele invloed van opdrachtgevers, uitvoerders en beneficiënten op selectie en inhoud van de geraadpleegde documentatie, respondenten, enz.
2.3.2	idem met betrekking tot de evaluatoren
2.4.1	uitvoering conform tijdplanning en begroting; oorzaken van afwijkingen
2.4.2	beschrijving - en verklaring - van het verloop van de evaluatie, inclusief noodzakelijk gebleken aanpassingen t.o.v. de oorspronkelijke opzet
2.5.1	zijn belanghebbenden in de gelegenheid gesteld commentaar op concept-bevindingen e.d. te leveren en is dergelijk commentaar door de onderzoekers verwerkt?
2.5.2	intensiteit van het toezicht op de opzet en/of uitvoering van het onderzoek door een "interne" (binnen de MFO[s]) of "externe" (buiten de MFO[s]) begeleidings- of stuurgroep o.i.d.
bruikbaarheid	
3.1.1	helderheid en begrijpelijkheid van bevindingen, conclusies en aanbevelingen
3.1.2	praktische uitvoerbaarheid van gedane aanbevelingen binnen het bereik van betrokken verantwoordelijken en het onderzochte werktelein
3.2.1	hoe volledig worden de, d.w.z. <u>alle</u> , onderzoeksvragen door de conclusies beantwoord?
3.2.2	worden positieve en negatieve verschillen tussen planning en realisatie van de onderzochte activiteiten niet alleen vastgesteld, maar ook tot hun oorzaken herleid?
3.3.1	is de samenvatting ondubbelzinnig en bevat zij de essentie van het rapport, op zijn minst de hoofdbevindingen?
3.3.2	niet alleen redactionele scheiding van conclusies en aanbevelingen in de tekst, maar ook <u>logische onderscheiden</u> van elkaar

¹⁹ SMART staat voor: Specifiek, Meetbaar, *Attainable* [Bereikbaar], Relevant en Tijdgebonden

www.euforic.org/iob

Ministerie van Buitenlandse Zaken
Postbus 20061
2500 EB Den Haag

www.minbuza.nl

ISSN 15166-3000

