

Praktijkevaluatie in de geestelijke gezondheidszorg: de modelcyclus

door A.A.J.M. Beenackers

Samenvatting

De belangrijkste ontwikkeling die zich de laatste jaren op evaluatiegebied heeft voorgedaan, is de aanzienlijke uitbreiding die het evaluatieterrein heeft ondergaan: voor wat de geestelijke gezondheidszorg betreft, heeft het evaluatieconcept zich daardoor verruimd van effectmeting tot beleidsevaluatie, waarbij het experimentele onderzoeksdesign vervangen werd door een longitudinale benadering. Tegelijkertijd werd in de Engelstalige literatuur de term 'evaluation research' vervangen door 'program evaluation'. Voor wat de geestelijke gezondheidszorg betreft, heeft dit laatste begrip thans betrekking op allerlei vormen van praktijkevaluatie; variërend van de evaluatie van een bepaalde behandelingsvorm tot de evaluatie van het overheidsbeleid. In dit artikel brengen we al deze vormen terug tot hetzelfde grondmodel: een cyclisch proces van exploratie, normering, operationalisatie, toetsing en bijsturing. Vervolgens werken we dit model verder uit, zodat het als richtsnoer kan dienen bij het opzetten van iedere vorm van praktijkevaluatie. Het model kan tevens een hulpmiddel zijn bij het analyseren van problemen die zich bij lopende evaluatieprocessen voordoen.

Inleiding

In de geestelijke gezondheidszorg (ggz) worden velerlei begrippen gehanteerd om de diverse vormen van praktijkevaluatie aan te duiden. Naast zeer algemene begrippen als 'kwaliteitsbewaking', 'kwaliteitscontrole' en 'gemeenschapscontrole', zijn dat vooral begrippen waarmee een bepaalde vorm van evaluatieonderzoek wordt aangeduid: 'programmaevaluatie', 'projectevaluatie', 'interne evaluatie', 'externe evaluatie', 'procesevaluatie', 'produktevaluatie', en dergelijke. Daarnaast wordt gebruik gemaakt van begrippen die verwijzen naar een specifieke techniek, procedure of doel; bij voorbeeld: 'goal attainment scaling', 'intercollegiale toetsing' en 'satisfactiemeting'.

Omdat de begrippen in betekenis variëren van zeer algemeen tot zeer specifiek, zijn op ieder concreet onderzoek verschillende begrippen tegelijk van toepassing. Een satisfactiemeting kan bij voorbeeld ook als evaluatieonderzoek gepresenteerd worden, of als een vorm van kwaliteitsbewaking. Maar ook binnen één bepaalde categorie van begrippen verwijst de terminologie naar steeds weer andere kenmerken van praktijkevaluatie. Zo verwijst 'kwaliteitscontrole' naar datgene

wat gecontroleerd wordt, terwijl 'gemeenschapscontrole' verwijst naar de controlerende instantie. Beide begrippen kunnen dus gelijktijdig van toepassing zijn. Hetzelfde geldt voor de begrippen waarmee een bepaalde vorm van evaluatie wordt aangeduid. Zo verwijzen de begrippen 'procesevaluatie' en 'produktevaluatie' naar een bepaald gedeelte van het praktijkveld, terwijl de begrippen 'interne evaluatie' en 'externe evaluatie' verwijzen naar de relatie tussen degene die evalueert en degene die geëvalueerd wordt. Paarsgewijs sluiten deze begrippen elkaar weliswaar uit, maar onderling zijn ze goed te combineren. Een extra begrip is dat de betekenis van een aantal van deze begrippen aan verandering onderhevig is, waarschijnlijk doordat het gebruik ervan nogal modegevoelig is. Een begrip dat aan populariteit wint, wordt in steeds meer situaties toegepast (en vice versa), waardoor het al snel een bredere betekenis krijgt. Andere begrippen krijgen daardoor een meer beperkte betekenis. Een populair begrip kan op die manier (tijdelijk) uitgroeien tot een containerbegrip, waarbij andere begrippen (tijdelijk) naar de achtergrond verdwijnen. Resumerend kunnen we stellen dat er ter typering van de verschillende vormen van praktijkevaluatie veel andersoortige begrippen in omloop zijn, die bovendien niet eenduidig gedefinieerd zijn en al evenmin consequent gehanteerd worden. Het is daarom niet gemakkelijk zich een duidelijk beeld te vormen van de essentie van praktijkevaluatie, en van de verschillen die er tussen de diverse varianten bestaan.

In dit artikel proberen we tussen de bomen door het bos weer zichtbaar te maken. We doen dat door alle vormen van praktijkevaluatie terug te brengen tot hetzelfde grondmodel. Dit grondmodel leiden we af uit het grondmodel dat voor theorie-evaluatie algemeen aanvaard is. We kiezen daarvoor de formulering van De Groot (1961). Volgens De Groot kan iedere vorm van theorie-evaluatie opgevat worden als een cyclisch proces van exploratie, inductie, deductie, toetsing en evaluatie. Voor praktijkevaluatie geldt hetzelfde. Er is evenwel één belangrijk verschil: bij theorie-evaluatie worden de ideeën (i.c. de desbetreffende theorie) systematisch aangepast aan de werkelijkheid, terwijl bij praktijkevaluatie de werkelijkheid (i.c. de desbetreffende praktijk) systematisch wordt aangepast aan de ideeën. Theorie-evaluatie en praktijkevaluatie zijn in dit opzicht dus twee tegengestelde ervaringsprocessen. Het grondmodel voor praktijkevaluatie ziet er daardoor uit als het spiegelbeeld van het grondmodel voor theorie-evaluatie. In dit artikel werken we dit grondmodel verder uit. We doen dit zodanig dat het resultaat beschouwd kan worden als een abstractie van allerlei vormen van praktijkevaluatie. Dit heeft enerzijds tot gevolg dat de verschillen tussen de diverse vormen van praktijkevaluatie teruggebracht worden tot accentverschillen, waardoor het gemakkelijker wordt zich een duidelijk beeld te vormen van de essentie van praktijkevaluatie; anderzijds krijgen we daardoor de beschikking over een model dat als richtsnoer kan dienen bij het opzetten van iedere vorm van

praktijkevaluatie. Het model kan tevens een hulpmiddel zijn bij het analyseren van problemen die zich bij lopende evaluatieprocessen voordoen.

Praktijkevaluatie versus theorie-evaluatie

Bij wetenschappelijk onderzoek worden meestal twee hoofddoelstellingen onderscheiden:

- (a) het vergroten van ons inzicht in de werkelijkheid; en
- (b) het aanpassen van de werkelijkheid aan onze inzichten.

Ons inzicht in de werkelijkheid wordt daarbij uitgedrukt in theorieën over de wijze waarop verschijnselen of gebeurtenissen met elkaar samenhangen. Het vergroten van ons inzicht in de werkelijkheid (doelstelling a) komt dus neer op het ontwerpen en verifiëren van relevante theorieën. Als we het omgekeerde doen – de werkelijkheid aanpassen aan onze inzichten –, ontstaat er een werkelijkheid die we gedeeltelijk zelf geconstrueerd hebben, en die we in de ggz als 'praktijk' kunnen aanduiden. Het aanpassen van de werkelijkheid aan onze inzichten (doelstelling b) komt dus neer op het ontwerpen en realiseren van door ons gewenste praktijksituaties. Bij onderzoek waarbij de nadruk ligt op doelstelling a, wordt nagegaan wat een bepaalde theorie waard is; bij onderzoek waarbij de nadruk ligt op doelstelling b, wordt nagegaan wat een bepaalde praktijk waard is. In het eerste geval kunnen we daarom spreken van theorie-evaluatie; in het tweede geval praktijkevaluatie. Zoals theorie-evaluatie gericht is op theorievorming, is praktijkevaluatie gericht op praktijkvorming.

Volgens De Groot (1961) is het verwerven van ervaring niet te begrijpen zonder de aanname van een cyclisch proces, dat in zijn meest algemene vorm bestaat uit 'waarnemen-proberen-evalueren'. Voor verschillende ervaringsprocessen kunnen uiteenlopende cyclusvormen bedacht worden. Al deze cyclusvormen moeten volgens De Groot gezien worden als uitdrukkingen van eenzelfde grondverschijnsel. Zonder de aanname van dit grondverschijnsel is geen enkel ervaringsproces verklaarbaar. Wetenschapsbeoefening is zo'n ervaringsproces. De door De Groot in zijn boek uitgewerkte cyclusvorm heeft betrekking op wat we zojuist 'theorie-evaluatie' genoemd hebben. Analoot aan deze uitwerking kunnen we voor praktijkevaluatie een soortgelijke cyclus ontwerpen, waarbij we bijna dezelfde formuleringen kunnen gebruiken.

We beginnen met de meest primitieve cyclusvorm: waarnemen-proberen-evalueren. Bij theorie-evaluatie kunnen we het 'proberen' omschrijven als 'nagaan of ons inzicht (de theorie) in overeenstemming is met de werkelijkheid'. Bij praktijkevaluatie kunnen we het 'proberen' omschrijven als 'nagaan of de werkelijkheid (de praktijk) in overeenstemming is met ons inzicht'. Dit 'proberen' voltrekt zich bij De Groot in een drietal fasen: inductie, deductie en toetsing. Bij prak-

tijkevaluatie kunnen we dezelfde driedeling maken: inductie, deductie en toetsing. We krijgen dan twee vergelijkbare cycli: zie figuur 1.

Figuur 1: Theorie-evaluatie versus praktijkevaluatie

	Theorie-evaluatie		Praktijkevaluatie
1.	Exploratie: het beschrijven van de verifiëren theorie	1.	Exploratie: het beschrijven van de beoogde praktijk
2.	Inductie: het formuleren van een aantal hypothesen, die in overeenstemming zijn met de te verifiëren theorie	2.	Inductie: het formuleren van een aantal normen, waaraan de beoogde praktijk moet voldoen
3.	Deductie: het afleiden van toetsbare voorspellingen uit de hypothesen	3.	Deductie: het afleiden van toetsbare criteria uit de normen
4.	Toetsing: het nagaan of de voorspellingen uitkomen	4.	Toetsing: het nagaan of aan de criteria wordt voldaan
5.	Evaluatie: het eventueel bijstellen van de theorie	5.	Evaluatie: het eventueel bijsturen van de praktijk

Theorie-evaluatie en praktijkevaluatie zijn aldus terug te brengen tot dezelfde grondvorm. Opvallend is dat de beide cycli in zekere zin elkaars spiegelbeeld zijn. Dat komt uiteraard doordat de desbetreffende ervaringsprocessen in bepaalde opzichten tegengesteld zijn aan elkaar: bij theorie-evaluatie worden de ideeën (de theorie) aangepast aan de werkelijkheid, terwijl bij praktijkevaluatie de werkelijkheid (de praktijk) wordt aangepast aan de ideeën. Zowel bij theorie-evaluatie als bij praktijkevaluatie begint men met een exploratieve fase. Deze exploratie levert een beschrijving op van de te verifiëren theorie c.q. de beoogde praktijk. In de volgende fase worden deze beschrijvingen aangescherpt: in het eerste geval in de vorm van hypothesen, en in het tweede geval in de vorm van normen. In beide gevallen spreken we van inductie. Een hypothese is een veronderstelling over een samenhang in de werkelijkheid. Door uit deze veronderstelling toetsbare voorspellingen af te leiden (deductie), kan door middel van toetsing nagegaan worden of deze veronderstelling juist is. Bij praktijkevaluatie hebben we te maken met normen: voorwaarden waaraan de praktijk moet voldoen. Door uit deze normen toetsbare criteria af te leiden (deductie), kan door middel van toetsing nagegaan worden of de praktijk aan de gestelde voorwaarden voldoet. Fase 3 is dus in beide gevallen de operationalisatiefase. Bij theorie-evaluatie bestaat fase 4 uit het toetsen van de voorspelling aan de werkelijkheid. Bij praktijkevaluatie zien we hier het spiegelbeeld: de werkelijkheid wordt getoetst aan het criterium. Fase 5 is in beide gevallen de evaluatiefase; de fase waarin bekeken wordt wat de betekenis is van de uitslag van de toetsing voor

de desbetreffende theorie c.q. praktijk. Bij theorie-evaluatie wordt hier de theorie geëvalueerd, terwijl bij praktijkevaluatie de praktijk geëvalueerd wordt. Vandaar dat de door ons gekozen termen 'theorie-evaluatie' en 'praktijkevaluatie' taalkundig correct zijn.

Bij theorie-evaluatie is het in principe zo dat als bij de toetsing blijkt dat de voorspelling niet uitkomt, de theorie wordt bijgesteld. En voor praktijkevaluatie geldt hetzelfde principe: als bij de toetsing blijkt dat niet aan het criterium wordt voldaan, wordt de praktijk bijgestuurd. Het komt uiteraard ook voor dat men achteraf tot de conclusie komt dat de voorspellingen niet goed geformuleerd waren, of dat de criteria niet realistisch waren, of dat de toetsing niet goed is uitgevoerd, en dergelijke. In dat geval kan men ook tot andere aanpassingen besluiten. Maar dat doet op zichzelf niets af aan het principe dat bij theorie-evaluatie steeds de theorie bijgesteld wordt, en bij praktijkevaluatie de praktijk.

Zowel bij theorie-evaluatie als bij praktijkevaluatie hebben we te maken met recurrerende cycli, waarbij telkens opnieuw de fasen 1 t/m 5 doorlopen worden. De laatste fasen van iedere cyclus gaan daarbij geleidelijk over in de eerste fasen van de volgende cyclus. In beginsel is het zo dat bij theorie-evaluatie iedere cyclus tot een zodanige bijstelling van de theorie moet leiden, dat bij de volgende toetsing de kans dat de voorspelling uitkomt groter is. Voor praktijkevaluatie geldt hetzelfde: in beginsel moet iedere cyclus tot een zodanige bijsturing van de praktijk leiden, dat bij de volgende toetsing de kans dat aan de criteria voldaan wordt groter is. In beide gevallen kan er een moment komen dat het evaluatieproces voltooid is, namelijk zodra (bij theorie-evaluatie) de voorspellingen uitkomen, c.q. (bij praktijkevaluatie) aan de criteria wordt voldaan. Edoch, bij theorie-evaluatie kan het stelsel van hypothesen steeds verder uitgebreid worden, waardoor een steeds groter gebied van de werkelijkheid wordt bestreken. Het vaststellen van de samenhang tussen een tweetal verschijnselen is op zichzelf namelijk niet zo interessant; het wordt pas interessant als we algemene samenhangen vinden die voor hele klassen van verschijnselen gelden (De Groot 1961). Voor praktijkevaluatie geldt in zekere zin hetzelfde. Het normeren van een beperkt aantal aspecten van een bepaalde praktijksituatie is op zichzelf niet zo interessant; het wordt pas interessant als we een stelsel onderling samenhangende normen vinden, die voor hele klassen van praktijksituaties gelden. Ook hier is het dus de bedoeling een steeds groter gebied van de werkelijkheid te bestrijken.

De structurele grondslag

Evaluatie van de praktijk betekent dat de gerealiseerde praktijk wordt vergeleken met de beoogde praktijk. De gegevens, die op de beoogde praktijk betrekking hebben, vormen *de normatieve grondslag* van een evaluatieproces; de gegevens die op de gerealiseerde praktijk betrek-

king hebben, vormen de *empirische grondslag* (Schnabel 1984). Evalueren betekent dus dat de empirische grondslag vergeleken wordt met de normatieve grondslag. De context waarbinnen zo'n vergelijking plaatsvindt, kan men beschouwen als *de structurele grondslag*. De structurele grondslag wordt vooral bepaald door de keuze van het evaluatieterrein en de rolverdeling tussen de verschillende instanties die bij het evaluatieproces betrokken zijn. Omdat de structurele grondslag geëxpliciteerd moet worden voordat men kan gaan evalueren, geven we in deze paragraaf een beschrijving van de grote lijnen die gezamenlijk het raamwerk vormen van de structurele grondslag, alvorens in de volgende paragraaf ons evaluatiemodel verder uit te werken.

Rolverdeling – Praktijkevaluatie is een vorm van terugkoppeling. Terugkoppeling houdt in dat van een bepaalde activiteit die achtereenvolgens is geprogrammeerd en uitgevoerd, het bereikte resultaat wordt vergeleken met het beoogde resultaat. Indien het bereikte resultaat niet gelijk is aan het beoogde resultaat, wordt de activiteit opnieuw geprogrammeerd en uitgevoerd, waarna vervolgens het bereikte resultaat weer vergeleken wordt met het beoogde resultaat. Enzovoorts; net zolang tot het bereikte resultaat gelijk is aan het beoogde resultaat. We hebben dan te maken met een cyclisch proces van programmering, uitvoering en terugkoppeling. Bij dit proces zijn in principe een drietal instanties betrokken: een instantie die het programma vaststelt: *de beleidsinstantie* (B); een instantie die het programma uitvoert: *de uitvoerende instantie* (U); en een instantie die het resultaat terugkoppelt van de uitvoerende instantie naar de beleidsinstantie: *de terugkoppelende instantie* (T). Ieder evaluatieproces is te beschouwen als een uitwerking van dit terugkoppelingsprincipe. Alle literatuur op dit gebied verwijst direct of indirect naar de 'feedback loop' uit de cybernetica (Wiener 1948).

Aan het in figuur 1 gepresenteerde evaluatiemodel ligt een tweede-ling ten grondslag tussen evaluerende instanties enerzijds (B en T), en geëvalueerde instantie anderzijds (U). Het in figuur 1 gepresenteerde model heeft namelijk uitsluitend betrekking op de evaluerende instanties: degene die de beoogde praktijk definieert (fase 1), de normen vaststelt (fase 2) en de praktijk bijstuurt (fase 5), vervult per definitie de rol van beleidsinstantie (B); degene die de normen operationaliseert (fase 3) en de toetsing uitvoert (fase 4), fungeert als terugkoppelende instantie (T).

In de ggz onderscheidt men gewoonlijk een vijftal beleidsniveaus: een behandelniveau (1), een afdelingsniveau (2), een instellingsniveau (3), een regionaal niveau (4) en een landelijk niveau (5). Ieder van deze beleidsniveaus kan in een evaluatieproces als beleidsinstantie (B) fungeren. Voor zover het beleid op de verschillende niveaus op elkaar afgestemd moet zijn, moeten uiteraard ook de normen die op de ver-

schillende niveaus gehanteerd worden met elkaar in overeenstemming zijn. Als alle niveaus gelijktijdig in één evaluatieproces zouden participeren, krijgen we een soort evaluatiepiramide (Mastboom 1984), waarbij de beleidslijnen van boven naar beneden lopen ($5 \rightarrow 1$), en de terugkoppellingslijnen van beneden naar boven ($1 \rightarrow 5$). De tussenliggende niveaus vervullen daarbij een dubbelrol: beleidsinstantie (B) ten opzichte van het lager gelegen beleidsniveau, en uitvoerende instantie (U) ten opzichte van het hogere beleidsniveau.

Als de beleidsinstantie ten opzichte van de uitvoerende instantie een externe organisatie is, spreekt men meestal van 'externe evaluatie'. In nagenoeg alle gevallen heeft dit betrekking op de beleidsniveaus 4 en 5 (m.n. overheid, inspectie en ziekenfonds – zowel regionaal (4) als landelijk (5)). Van interne evaluatie is sprake als beleidsinstantie (B) en uitvoerende instantie (U) tot dezelfde organisatie behoren. Dit speelt zich meestal af op de niveaus 1 (behandelingsniveau), 2 (afdelingsniveau) en 3 (instellingsniveau). Maar ook op de niveaus 4 en 5 kan sprake zijn van interne evaluatie, met name als regionale (4) en landelijke (5) koepelorganisaties als beleidsinstantie fungeren.

Evaluatieterrein – Alle activiteiten die als geestelijke gezondheidszorg aangemerkt kunnen worden, kunnen tot het praktijkveld van de ggz gerekend worden. Uiteraard is het niet mogelijk om in één evaluatieproces de hele ggz te evalueren. Ieder evaluatieproces bestrijkt daarom slechts een gedeelte van het totale veld, bij voorbeeld een bepaalde categorie behandelingen of een bepaald type hulpverlenende instellingen. In het eerste geval fungeren behandelingskenmerken als indelingscriterium, in het tweede geval instituutskennmerken. Meestal hebben we in de praktijk te maken met een combinatie van indelingscriteria, bij voorbeeld psychotherapieën (behandelingskenmerk), uitgevoerd door regionale instellingen voor ambulante geestelijke gezondheidszorg (instituutskennmerk). We hebben dan het evaluatieterrein afgebakend. Indien we een bepaald evaluatieterrein kiezen en benoemen, kunnen we in ons evaluatiemodel het woord 'praktijk' vervangen door de naam die we aan het gekozen evaluatieterrein gegeven hebben, zonder dat dit verder consequenties heeft. Stel dat we bij een $N = 1$ studie als evaluatieterrein kiezen: 'de behandeling van patiënt A'. 'Praktijkevaluatie' wordt dan: 'de evaluatie van de behandeling van patiënt A'. En fase 1: 'het beschrijven van de beoogde behandeling van patiënt A'. Fase 2: 'het formuleren van een aantal normen waaraan de behandeling van patiënt A moet voldoen'. Enzovoorts: zie figuur 1. Aldus kan het woord 'praktijk' vervangen worden door iedere omschrijving die een nadere specificatie van het evaluatieterrein inhoudt. En vaak wordt hierbij volstaan met één enkel woord, bij voorbeeld: 'programmaevaluatie', 'projectevaluatie', 'procesevaluatie' en 'produktevaluatie'.

Bij de specificatie van het evaluatieterrein kan het zinvol zijn onder-

scheid te maken tussen doel en middel, waarbij het middel systematisch gemanipuleerd wordt teneinde het gewenste doel te bereiken. De normering (fase 2) heeft dan betrekking op het doel, en de bijsturing (fase 5) op het middel. Zo kan men bij voorbeeld een bepaalde behandelingsvorm als uitgangspunt kiezen, en de samenstelling van het patiëntenbestand daar systematisch bij aanpassen. Of omgekeerd: uitgaan van de patiëntenpopulatie die zich aanbiedt, en de behandelingsvormen daar systematisch bij aanpassen. Vooral bij grote instellingen, die uit verschillende afdelingen bestaan, is het voorstelbaar dat men op instellingsniveau voor de laatste variant kiest, terwijl men op afdelingsniveau meer de nadruk legt op de eerste variant. Het onderscheid tussen doel en middel hangt aldus direct samen met het evaluatieterrein dat men gekozen heeft: de instelling als geheel, of een bepaalde afdeling daarvan. En het evaluatieterrein hangt weer samen met het evaluatieniveau: niveau 3, of niveau 2. Vandaar dat ook het onderscheid tussen doel en middel tot de structurele grondslag gerekend kan worden.

De modelcyclus

Zoals we in figuur 1 gezien hebben, is iedere vorm van praktijkevaluatie te beschouwen als een cyclisch proces van exploratie, normering, operationalisatie, toetsing en bijsturing. Iedere evaluatiecyclus die volgens dit grondpatroon is opgebouwd, beschouwen we als een modelcyclus. In deze paragraaf werken we de modelcyclus verder uit. We bespreken daartoe iedere fase afzonderlijk.

Fase 1: exploratie – De exploratie dient een beschrijving op te leveren van de beoogde praktijk. Aan deze beschrijving worden vervolgens, in fase 2, de normen ontleend waaraan de praktijk op een nader te bepalen tijdstip moet voldoen. Een beschrijving van de beoogde praktijk valt uiteen in een beschrijving van de te realiseren doeleinden en de daartoe aan te wenden middelen. De normen (fase 2) en de criteria (fase 3) worden op de doelstellingen gebaseerd, terwijl de bijsturing (fase 5) betrekking heeft op de aan te wenden middelen. De relatie tussen de wijze waarop de middelen worden aangewend en de mate waarin de doelstellingen worden gerealiseerd, blijkt uit fase 4 (de toetsing). De toetsing stelt ons dus in staat om terug te schakelen van 'doel' naar 'middel'.

Om op realistische wijze aan te kunnen geven hoe de praktijk er in de toekomst uit moet zien, moet men een tamelijk goed inzicht hebben in de bestaande praktijk. Indien dit inzicht ontbreekt, moet het eerst verworven worden. Eén van de mogelijkheden daartoe is het doen van een uitgebreid exploratief onderzoek. Pas als de exploratie voltooid is, worden er normen geformuleerd. En pas als de normen vastgesteld zijn, worden deze geoperationaliseerd. Enzovoorts. Zowel

de doelstellingen als de normen en de criteria hebben dan een tamelijk definitief karakter, dat wil zeggen dat ze tijdens het evaluatieproces nauwelijks zullen veranderen. De opeenvolgende cycli bestaan dan voornamelijk uit een zich steeds herhalende toetsing (fase 4) en bijsturing (fase 5), waarbij de wijze waarop de middelen aangewend worden consequent wordt aangepast aan de te realiseren doeleinden – een proces dat in principe wordt voortgezet tot de praktijk aan de gestelde criteria voldoet.

In plaats van een exploratief onderzoek te doen in fase 1, kan men ook de cyclus als geheel een exploratief karakter geven. Men neemt dan genoeg met een voorlopige beschrijving van de beoogde praktijk. Zo'n beschrijving kan zich bij voorbeeld beperken tot een aantal voorlopige doelstellingen. In fase 2 worden dan voorlopige normen geformuleerd, waaruit in fase 3 voorlopige criteria afgeleid worden. Bij de toetsing dient vervolgens voldoende informatie beschikbaar te komen om op grond daarvan een meer nauwkeurige beschrijving te kunnen geven van de beoogde praktijk. En zo'n meer nauwkeurige beschrijving stelt ons weer in staat de doelstellingen, normen en criteria een meer definitief karakter te geven. De opeenvolgende cycli omvatten dan alle fasen, waarbij de fasen 4 en 5 van iedere cyclus geleidelijk overgaan in de fasen 1 en 2 van de volgende cyclus. Het sterk exploratieve karakter van de eerste cycli maakt daarbij geleidelijk plaats voor een meer toetsend karakter bij navolgende cycli. Ook dit proces wordt in principe voortgezet tot de praktijk aan de gestelde criteria voldoet. Maar anders dan bij een uitgebreide exploratie vooraf, staan de criteria niet reeds bij de eerste cyclus vast, maar krijgen ze geleidelijk hun definitieve vorm.

De twee zojuist genoemde exploratiemogelijkheden zijn te beschouwen als twee uitersten, waartussen diverse gradaties mogelijk zijn. En het is niet noodzakelijk om voor ieder criterium dezelfde vorm van exploratie te kiezen: men kan in één evaluatieproces, waarbij meerdere criteria tegelijk gehanteerd worden, starten met een combinatie van min of meer definitieve criteria en min of meer voorlopige criteria. Vaak is zelfs vooraf niet eens te zeggen tot welk van de twee categorieën een bepaald criterium behoort. Dat blijkt dan pas achteraf, na één of meerdere cycli.

Ter adstructie van de verschillende exploratiemogelijkheden verwijzen we naar de evaluatie van de crisiscentra van Overvecht (1974-1978) en Utrecht (1978-1982). Bij de start van het eerste experiment was geen enkel criterium voorhanden. Aan de hand van één voorlopige doelstelling (namelijk: 'crisiscliënten moeten geholpen worden') was na één maand de eerste cyclus voltooid. Door dit proces steeds te herhalen, en door daarbij de voorlopige doelstelling steeds verder uit te werken, ontstond als vanzelf de zojuist geschetste ontwikkeling van exploratie naar toetsing (Beenackers 1982). Bij de aanvang van het tweede experiment ging het eerste experiment fungeren als een uitge-

breide exploratie vooraf, dat wil zeggen dat bij het eerste experiment voldoende inzicht verworven was om bij het tweede experiment vooraf de normen definitief vast te leggen. De opeenvolgende cycli kregen toen het karakter van een zich steeds herhalende toetsing (fase 4) en bijsturing (fase 5), waarbij de wijze waarop de middelen aangewend werden consequent werd aangepast aan de te realiseren doeleinden (Beenackers 1984).

Fase 2: normering – Een norm is een uitspraak die op een bepaald moment voor een bepaalde praktijksituatie óf waar is, óf niet waar is; en waarbij vooraf is bepaald dat de desbetreffende praktijksituatie daarmee in het ene geval positief wordt gewaardeerd, en in het andere geval negatief. Het begrip 'norm' heeft hier dus de betekenis van een beoordelingsmaatstaf, waarbij er in principe maar twee mogelijkheden zijn: er wordt wél aan voldaan, of er wordt níet aan voldaan. Normen moeten enerzijds afgeleid zijn van de (in fase 1 beschreven) beoogde praktijk, en anderzijds de mogelijkheid bieden er (in fase 3) criteria uit af te leiden, die ons in staat stellen eenduidig vast te stellen of op een bepaald moment voor een bepaalde praktijksituatie de uitspraak waar is, of niet waar is. De eerste eis is inhoudelijk van aard, de tweede formeel. Voor wat de hoogte van de norm betreft, moet men bedenken dat evaluatie pas zinvol is als niet vooraf te zeggen valt of de praktijk er wél of niet aan zal voldoen. Beide uitkomstmogelijkheden moeten dus reëel zijn. Als de uitkomst vooraf bekend is, heeft toetsen immers geen zin. Daarnaast is het zo dat alleen van normen die iets hoger gesteld zijn dan het actuele prestatieniveau, een prestatieverbeterende invloed verwacht mag worden. De toetsing moet dan uitwijzen of zo'n prestatieverbetering inderdaad heeft plaatsgevonden.

Fase 3: operationalisatie – Normen zijn slechts zelden zo geformuleerd, dat eenduidig vastgesteld kan worden of de praktijk er wél of niet aan voldoet. Ze moeten daartoe meestal eerst geoperationaliseerd worden; dat wil zeggen dat exact aangegeven moet worden hoe vastgesteld moet worden of de praktijk er al dan niet mee in overeenstemming is. Operationaliseren impliceert meestal een tweetal stappen: eerst worden de normen geherformuleerd, zodat we uitspraken krijgen waaraan de praktijk wél eenduidig getoetst kan worden, en vervolgens wordt aangegeven hoe de toetsing uitgevoerd moet worden. Ingeval we bij voorbeeld als norm gesteld hebben 'de hulpverlening is adequaat', zou dit in eerste instantie kunnen worden 'de hulpverleningsdoelstellingen worden gerealiseerd', en in tweede instantie 'tenminste 80% van de goal attainment-scores is positief'. We hebben de norm dan omgezet in een toetsbaar criterium. Bij de toetsing (fase 4) neemt het criterium als het ware de plaats in van de norm. De hulpverlening wordt dus adequaat geacht, als de goal attainment-scores de vereiste waarden hebben. De stap van norm naar criterium betekent objectivi-

teitswinst; dat wil zeggen dat de praktijk zich objectiever laat beoordelen aan de hand van het criterium dan aan de hand van de norm. Maar operationalisatie levert meestal behalve winst ook verlies op. Het verlies treedt op bij de herformulering, omdat het criterium als regel een verbijzondering is van de norm, waarbij het criterium in vergelijking tot de norm aan relevantie heeft ingeboet. 'Objectiviteitswinst' moet dus vaak betaald worden met 'relevantieverlies'. Men kan daarbij voor de volgende keuze gesteld worden: óf de norm net zolang herformuleren tot er een objectief criterium ontstaat, waarbij men accepteert dat dit criterium minder relevant is dan de oorspronkelijke norm, óf zo weinig mogelijk relevantie prijsgeven, waarbij men accepteert dat de toetsing enigszins subjectief is. In het eerste geval kiest men voor betrouwbaarheid boven validiteit, in het tweede geval voor validiteit boven betrouwbaarheid. In zijn algemeenheid kan gesteld worden dat validiteit een eerste vereiste is, en betrouwbaarheid een tweede (De Groot 1961). Vandaar dat normen eerst geherformuleerd worden, en pas daarna van een toetsingsvoorschrift worden voorzien; eerst wordt het relevantieprobleem opgelost, en pas daarna het objectiviteitsprobleem.

Fase 4: toetsing – Onder 'toetsing' verstaan we het zo objectief mogelijk verzamelen en bewerken van gegevens, zodat eenduidig komt vast te staan of de praktijk wél of niet aan de gestelde criteria voldoet. Bij het verzamelen van de gegevens heeft men de keuze tussen het verzamelen van een gedeelte van de daarvoor in aanmerking komende gegevens, en het verzamelen van alle gegevens. Ingeval er met deelverzamelingen wordt gewerkt, worden er in principe alleen tijdens de toetsingsfase gegevens verzameld. Complete verzamelingen worden meestal aangelegd met behulp van een registratiesysteem, dat dagelijks bijgehouden moet worden. Het bekendste voorbeeld hiervan is het Amerikaanse Management Information System (MIS), alhoewel dit in feite slechts een verzamelnaam is voor allerlei informatiesystemen (Schnabel 1984), zoals we die inmiddels ook in Nederland kennen.

Bij het continu verzamelen van gegevens doet zich al snel de wet van de afnemende meeropbrengst gelden, omdat over het algemeen deelverzamelingen dezelfde informatie opleveren als complete verzamelingen. Het continu verzamelen van gegevens is alleen maar zinvol als we te maken hebben met korte cycli, die elkaar in snel tempo opvolgen – bij voorbeeld ieder kwartaal een nieuwe cyclus. Zo'n hoge frequentie is alleen dan noodzakelijk als de evaluatie betrekking heeft op verschijnselen die een sterk fluctuerend karakter hebben, en waarbij de fluctuaties die zich daarbij voordoen zodanig van aard zijn dat ze direct vertaald moeten worden in beleidsmaatregelen. Onze indruk is dat dit veel minder vaak het geval is dan het aantal permanente registratiesystemen doet vermoeden.

Behalve dat continu registreren tijdrovend is, kan het ook verstarrend werken. Over het algemeen kost het namelijk veel moeite om zo'n systeem ingevoerd te krijgen, waardoor men zich wel drie keer zal bedenken alvorens er weer veranderingen in aan te brengen. Men loopt dan het risico dat de verzamelde gegevens gaan fungeren als het uitgangspunt van het evaluatieproces, terwijl ze daarvan in feite het sluitstuk vormen. Daarnaast is het zo dat het werken met permanente registratiesystemen het gebruik van onderling complementaire methoden van dataverzameling in de weg kan staan. En juist het tegelijkertijd gebruiken van verschillende manieren van dataverzameling, zoals interviews, observatie en documentenanalyse, leidt tot de grootst mogelijke validiteit en betrouwbaarheid (Van der Zwaan 1981).

Fase 5: bijsturing – In beginsel moet ten aanzien van ieder criterium, waaraan bij de toetsing niet blijkt te worden voldaan, bijsturing van de praktijk plaatsvinden, zodat na verloop van tijd alsnog aan het criterium wordt voldaan. Indien de criteria zijn gebaseerd op de doelstellingen, worden in de bijsturingsfase de middelen die ter realisering daarvan worden aangewend, opnieuw ter discussie gesteld. De middelen mogen daarom niet te stringent genormeerd zijn. Dit, omdat er anders te weinig bijsturingsruimte beschikbaar is. Wie bij voorbeeld het hulpverleningsprodukt normeert, moet het hulpverleningsproces zo veel mogelijk vrij laten. Aldus kan het proces systematisch aangepast worden aan het te realiseren produkt. Het effect van de genomen bijsturingsmaatregelen blijkt pas bij fase 4 van de volgende cyclus, wanneer de praktijk opnieuw wordt getoetst. Uiteraard kan men ook tot de conclusie komen dat de doelstellingen, de normen of de criteria moeten worden bijgesteld. Maar ook dit wordt gerekend tot de volgende cyclus.

Uitleiding

De belangrijkste ontwikkeling die zich de laatste jaren op evaluatiegebied heeft voorgedaan, is de aanzienlijke uitbreiding die het evaluatieterrein heeft ondergaan: voor wat de geestelijke gezondheidszorg betreft, heeft het evaluatieconcept zich daardoor verruimd van effectmeting tot beleidsevaluatie, waarbij het experimentele onderzoeksdesign vervangen werd door een longitudinale benadering. Tegelijkertijd zien we dat de term 'evaluatieonderzoek' vervangen wordt door begrippen die op allerlei vormen van evaluatie van toepassing zijn; variërend van de evaluatie van een bepaalde behandelingsvorm tot de evaluatie van het overheidsbeleid. In dit artikel hebben we daarvoor de term 'praktijkevaluatie' gereserveerd: een containerbegrip waarin het woord 'praktijk' naar believen vervangen kan worden door een woord dat een meer specifieke betekenis heeft.

De sterke uitbreiding van het evaluatieterrein heeft geleid tot een door Grünwald & Kraan (1985) als een 'jungle' getypeerde veelheid aan evaluatievormen – zowel in de evaluatiepraktijk als in de meer theoretisch gerichte evaluatieliteratuur. Helaas kan ook de literatuur zelf als een 'jungle' getypeerd worden, omdat – zoals we al in de inleiding gesteld hebben – er ter aanduiding van de diverse evaluatievormen veel andersoortige begrippen in omloop zijn, die bovendien niet eenduidig gedefinieerd zijn en al evenmin consequent gehanteerd worden. In dit artikel hebben we getracht het evaluatieproces inzichtelijk te maken door de veelheid aan evaluatievormen terug te brengen tot een eenheid. We hebben daartoe een evaluatiemodel gepresenteerd, dat beschouwd kan worden als een abstractie van de meest ideale versies van allerlei vormen van praktijkevaluatie. Dit stelt ons in staat in een volgend artikel (Beenackers 1987) ook de evaluatieliteratuur toegankelijk te maken. We presenteren daartoe een typologie van evaluatieprocessen die niet beantwoorden aan het model dat we zojuist uiteengezet hebben. Het model zelf gaat dan als norm fungeren, en de evaluatieprocessen die niet aan deze norm voldoen, worden als minder adequaat beschouwd. Aan de hand van voorbeelden laten we dan zien hoe zo'n minder adequaat evaluatieproces getransformeerd kan worden in een meer adequaat proces.

Literatuur

- Beenackers, A.A.J.M. (1982), Project-evaluatie in de geestelijke gezondheidszorg: het Utrechtse model. *Gezondheid & Samenleving*, 3, 251-259.
- Beenackers, A.A.J.M. (1984), Het bestrijden van discriminerende hulpverlening, door toepassing van projectevaluatie. *Tijdschrift voor Psychiatrie*, 26, 42-57.
- Beenackers, A.A.J.M. (1987), Praktijkevaluatie in de geestelijke gezondheidszorg: onvolledige en afwijkende cyclusvormen. (Verschijnt in het *Tijdschrift voor Psychiatrie*.)
- Grünwald-Schindl, C.A., en A. Kraan-Jetten (1985), Evaluatie van en voor overheidsbeleid: een overzicht van vormen en benaderingen. *Beleid en Maatschappij*, 12, 128-138.
- Groot, A.D. de (1961), *Methodologie; grondslagen van onderzoek en denken in de gedragswetenschappen*. Mouton, 's-Gravenhage.
- Mastboom, J.C.M. (1984), Programma-evaluatie: voorwaarden en mogelijkheden. In: Schippers, Sijben, Schaap en Mastboom (red.), *Programma-evaluatie in de geestelijke gezondheidszorg*. Swets & Zeitlinger, Lisse, p. 48-57.
- Schnabel, P. (1984), Onderzoek en ordening in de geestelijke gezondheidszorg op regionaal niveau. *Kwartaalschrift Leiding & Organisatie in de Gezondheidszorg*, 1, 100-130.
- Wiener, N. (1948), *Cybernetics*. Wiley, New York.
- Zwaan, A.H. van der (1981), Evaluatie-onderzoek in organisaties: een belang voor de gezondheidszorg. *Gezondheid & Samenleving*, 2, 139-152.

Schrijver is psycholoog; tijdelijk werkzaam bij de RIAGG Stad Utrecht. Adres: Coornhertstraat 16-16^{bis}, 3521 XH Utrecht.