

Integraal erkend

Naar een afstemming erkenning jeugdinterventies

Tom van Yperen

Nederlands Jeugdinstituut / NJi

Juni 2007

©2007 Nederlands Jeugdinstituut / NJi

Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt door middel van druk, fotokopie, microfilm of op enige andere wijze zonder voorafgaande schriftelijke toestemming.

Auteur

Tom van Yperen

Informatie over dit document is te verkrijgen bij:

Nederlands Jeugdinstituut
Postbus 19221
3501 DE Utrecht
Telefoon (030) 230 65 61
E-mail: t.vanyperen@nji.nl

Inhoudsopgave

1. Inleiding	2
2. Uitgangspunt: onderzoek en praktijk	3
2.1 Twee routes	3
2.2 Wisselwerking bevorderen.....	3
2.3 Aansprekende criteria	4
3. Gehanteerde criteria in drie landelijke trajecten.....	5
3.1 Databank Effectieve Jeugdinterventies	5
3.2 Landelijke Erkenning Jeugdinterventies.....	5
3.3 Erkenningscommissie Gedragsinterventies Justitie	6
3.4 Discussie	7
4. Andere systemen	8
4.1 Uitputtende criteria?	8
4.2 Mogelijkheden van effectonderzoek.....	8
4.3 De gehanteerde normen.....	9
4.4 Conclusie.....	9
5. Een integraal kader.....	11
5.1 Ontwikkelingsniveau 0	11
5.2 Ontwikkelingsniveau I: in theorie effectief	11
5.3 Ontwikkelingsniveau II: bewezen effectief	13
5.4 Ontwikkelingsniveau III: kosteneffectief.....	16
5.5 Samenvattend model.....	16
6. Afspraken Erkenning Jeugdinterventies voor 2007	18
Bijlagen	19
Bijlage 1. Criteria Databank effectieve Jeugdinterventies.....	20
Bijlage 2. Criteria Preffi.....	22
Bijlage 3. Criteria Werkgroep Erkenning Jeugdinterventies	24
Bijlage 4. Criteria Erkenningscommissie Gedragsinterventies Justitie	27
Bijlage 5. Standards of evidence volgens Flay et al.....	31
Bijlage 6. Classificaties kwaliteit van onderzoek	34
Bijlage 7. Classificaties sterkte effect.....	37
Bijlage 8. Overzicht erkenning programma's in buitenlandse databanken	38

1. Inleiding

In de jeugdsector¹ is een ontwikkeling op gang gekomen waarin interventies op kwaliteit worden getoetst. Deze ontwikkeling hangt samen met een toenemend streven naar meer ‘evidence-based’ handelen: beroepsbeoefenaren baseren hun activiteiten op voorhandenzijnde kennis over wat het beste werkt in welke situatie. Doel van de toetsing op kwaliteit is in dat kader tweeledig:

- Het leidt ertoe dat het kaf van het koren wordt gescheiden. Interventies die slecht zijn onderbouwd, of blijkens onderzoek zelfs averechts kunnen uitpakken, vinden geen toepassing meer ten gunste van interventies die goed zijn onderbouwd en waarbij eventueel onderzoek laat zien dat er goede resultaten mee worden geboekt.
- Het stimuleert dat interventies goed worden onderbouwd en dat er onderzoek plaatsvindt naar de effectiviteit. Kwaliteitstoetsing levert een soort ‘opwaartse druk’ op in de sector: om te voorkomen dat een interventie wordt afgekeurd, is men bereid meer te investeren in theorievorming en onderzoek. Dat moet uiteindelijk de kwaliteitsontwikkeling in de hele sector ten goede komen.

Voor de beoordeling van de interventies bestaan uiteenlopende sets van criteria. Een oriëntatie op de literatuur en op databanken waarin informatie over interventies is opgenomen, laat zien dat er voornamelijk geen consensus bestaat over welke criteria van belang zijn. Deels kan dat te maken hebben met het feit dat de jeugdsector bestaat uit nogal uiteenlopende deelsectoren, met elk eigen tradities, wetenschappelijke maatstaven en standen van ontwikkeling. Daar staat tegenover dat vanuit de noodzaak samenwerking in de sector te bevorderen het van belang is te weten wat minimaal van de kwaliteit van interventies verwacht mag worden, en welke criteria voor de verdere ontwikkeling van de interventies van belang zijn.

In drie trajecten, de Databank Effectieve Jeugdinterventies, Erkenning Jeugdinterventies en Erkenning gedragsinterventies Justitie, zijn afgelopen jaren pogingen ondernomen de minimale kwaliteitscriteria te verhelderen. In deze notitie biedt een korte bespreking van deze criteria en plaatsen die tegen de achtergrond van beoordelingskaders die in de literatuur en bij internationaal bekende databanken te vinden zijn. Eerst komt het uitgangspunt aan bod waar de genoemde trajecten door gekenmerkt worden. Dan volgt een beknopte bespreking van de criteria die de trajecten bieden, en welke vraagstukken die oproepen. Tot slot volgt een voorstel voor een integraal kader waarbinnen uiteenlopende benaderingen te plaatsen zijn.

¹ Onder ‘jeugdsector’ verstaan we hier het geheel van professionele activiteiten in de jeugdzorg en psychosociale/pedagogische preventie, jeugdgezondheidszorg, preventie met betrekking tot jeugdgezondheid en gezondheidsbevordering, ontwikkelingsstimulering, onderwijs en jeugdwelzijn.

2. Uitgangspunt: onderzoek en praktijk

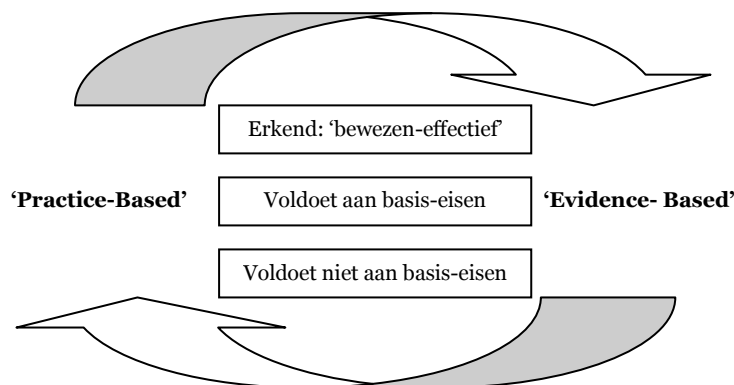
2.1 Twee routes

Interventies die bij de Databank Effectieve Jeugdinterventies, Erkenning Jeugdinterventies en Erkenning gedragsinterventies Justitie ter beoordeling worden aangeboden, komen grofweg van twee kanten.

- Enerzijds gaat het om interventies die in wetenschappelijke centra en/of in kleine ‘pilots’ in de praktijk ontwikkeld worden, op effectiviteit worden onderzocht en – bij gebleken kwaliteit - vervolgens breder worden geïmplementeerd. De implementatie gebeurt vaak onder het motto van het gebruik van ‘evidence-based’ (‘bewezen-effectieve’) interventies.
- Anderzijds worden er werkwijzen ingebracht die – soms op brede schaal - in de praktijk zijn gegroeid, verder zijn geëxpliciteerd en nader worden onderzocht op effectiviteit. Dit wordt ook wel aangeduid als de verdere ontwikkeling van ‘practice-based’ (‘in de praktijk gewortelde’) interventies.

De twee routes sluiten elkaar niet uit. In de praktijk gewortelde interventies kunnen zo zijn doorontwikkeld, dat ze een bewezen-effectieve karakter gaan krijgen en als zodanig verder worden verspreid. Ervaringen met bewezen-effectief programma’s kunnen leiden tot aanscherpingen in de praktijk die de effectiviteit verder ten goede kunnen komen. Als dergelijke verbeteringen van de effectiviteit in nieuw onderzoek worden bevestigd, is er feitelijk wederom sprake van een bewezen-effectief programma etc. Anders gezegd, in de huidige praktijk spelen zowel de implementatie van bewezen-effectieve interventies als ook het onderbouwen en empirisch toetsen van in de praktijk gewortelde programma’s een rol (zie ook figuur 1).

Figuur 1. Wisselwerking implementatie evidence-based en ontwikkeling practice-based interventies



Deze wisselwerking heeft voor de Erkenningstrajecten twee belangrijke consequenties: ten eerste moet het de geschetste wisselwerking bevorderen en ten tweede moet de werkwijze en de functie van de erkenning voor de betrokken partijen (wetenschappers en aanbieders) goed te snappen zijn.

2.2 Wisselwerking bevorderen

De erkenning van interventies moet ertoe leiden dat het zowel de implementatie van ‘evidence-based’ interventies als de onderbouwing van en het onderzoek naar in de praktijk geworteld aanbod bevordert. De ervaring van de afgelopen jaren leert dat het weinig productief is om in dat verband de erkenning uitsluitend te richten op interventies die volgens de hoogste maatstaven als bewezen-effectief mogen gelden. Een dergelijke benadering leidt ertoe dat er slechts weinig interventies kunnen

worden erkend. Op zich is die spaarzame erkenning nog geen probleem; het weerspiegelt in feite op een adequate manier de stand van de ontwikkeling in de sector. Het probleem is eerder dat – indien de erkenning zich uitsluitend richt op de hoogste maatstaven - er in de bulk van niet bewezen-effectieve programma's geen gedegen differentiatie komt in goede en minder goede interventies. Een dergelijke differentiatie is van belang om onderzoeksinvesteringen te kunnen richten op ten minste die interventies die aan een aantal basisvereisten voldoen. Interventies die daar niet aan voldoen zouden eerst doorontwikkeld moeten worden, alvorens daar onderzoek op te richten.

In de erkenningstrajecten is het daarom van belang onderscheid te maken in de verschillende ontwikkelingsstadia van interventies: interventies die niet en wel aan een aantal basisvereisten voldoen en interventies die daarbovenop als bewezen-effectief zijn aan te merken. Figuur 1 geeft dit in hoofdlijnen weer.

2.3 Aansprekende criteria

Bij de erkenning van interventies spelen wetenschappelijke criteria een belangrijke rol. Hierop zijn de eisen ten aanzien van theoretische onderbouwing en empirische toetsing gebaseerd. De ervaring leert dat dit aanleiding geeft tot nog een ander aandachtspunt: de gehanteerde criteria moeten niet alleen voor wetenschappers, maar ook voor aanbieders te volgen zijn. Is dat niet het geval, dan wordt de erkenning voor aanbieders al gauw een 'ver-van-mijn-bed-show'. Een dergelijke distantie kan verhinderen dat aanbieders zich betrokken voelen bij het belang van de erkenning en zich ook verantwoordelijk voelen bij te dragen aan de goede implementatie en/of verdere ontwikkeling van de interventie. Met andere woorden: de erkenning moet geen onderonsje zijn van wetenschappers, maar een gezamenlijk traject van verschillende belanghebbenden in de sector.

Dit betekent concreet dat de gehanteerde criteria voor de erkenning zo beperkt en inzichtelijk mogelijk moeten zijn. Het moet om de essentie gaan, en niet om het meest omvattende ideaal. De criteria moeten bij wijze van spreken aan iedereen uit te leggen zijn.

3. Gehanteerde criteria in drie landelijke trajecten

Richten we nu de blik op de inhoud van de gehanteerde criteria in de drie landelijke trajecten, dan zien we het volgende.

3.1 Databank Effectieve Jeugdinterventies

De *Databank Effectieve Jeugdinterventies* hanteert een set van criteria die is opgesteld aan de hand van uitgevoerde beoordelingstrajecten in de Voor- en Vroegschoolse Educatie (VVE), pedagogische preventie en jeugdzorg (zie bijlage 1). De databank maakt op basis van die criteria onderscheid tussen:

- interventies die niet-veelbelovend zijn;
- interventies die aan basiscriteria voldoen en daarmee veelbelovend zijn ('in theorie effectief');
- interventies die veelbelovend zijn en waar bovendien onderzoek positieve resultaten laat zien. Het onderzoek is aan de hand van een zogeheten 'sterrensysteem' op een aantal basale kenmerken in kwaliteit gedifferentieerd. Daarnaast is voor de aanduiding van de aard en omvang van de positieve resultaten een voorlopige classificatie ontwikkeld. Er zijn nog geen criteria geformuleerd voor de vraag wanneer een interventie als 'evidence-based' is aan te duiden. De kwaliteit en de uitkomsten van het onderzoek worden beschrijvend weergegeven.

De criteria zijn uitgewerkt in werkbladen waarmee aanbieders in principe – met ondersteuning van het NJi – hun interventie kunnen beschrijven en analyseren. NJi-medewerkers vatten de analyse samen in een voorlopig kwaliteitsoordeel over de interventie. Zij doen dit in dialoog met de aanbieders: het voorlopig kwaliteitsoordeel geeft meteen suggesties over de aspecten waarop interventies te verbeteren zijn. Onafhankelijke panels doen vervolgens definitieve uitspraken over de kwalificatie van de beschreven interventies. Ook treden deze panels op wanneer aanbieders en NJi-medewerkers van mening verschillen over het voorlopige kwaliteitsoordeel. De erkende interventies, alsook de interventies die niet zijn erkend of door het NJi in samenspraak met de aanbieder niet aan de panels zijn voorgelegd zijn gepubliceerd op internet (www.jeugdinterventies.nl).

De databank maakt om de twee jaar een update van de stand van zaken omtrent een interventie.

3.2 Landelijke Erkenning Jeugdinterventies

De *Werkgroep Erkenning Jeugdinterventies* heeft op basis van met name de criteria van de Databank Effectieve Jeugdinterventies en de zogeheten Preffi (zie bijlage 2) de set van criteria verder gedifferentieerd. Zie bijlage 3 voor het resultaat.

Bij de criteria is een onderscheid gemaakt in vier kwaliteitskenmerken van de interventies, die samenhangen met de predikaten die de commissie afgeeft:

- *Minimale kwaliteit van de interventies.* Interventies die niet voldoen aan minimale vereisten ten aanzien van de onderbouwing, beschrijving en kwaliteitsbewaking (zie volgende niveau) worden als 'niet erkend' aangeduid.
- *De kwaliteit van de theoretische onderbouwing, de mate van volledigheid van de beschrijving (doelgroep, doel en activiteiten), de randvoorwaarden en eisen voor kwaliteitsbewaking.* Interventies die aan een minimale set van kwaliteitscriteria voor dit kenmerk voldoen, ontvangen het predikaat 'in theorie effectief', doelend op de veronderstelling dat – gegeven de kwaliteit – de interventie effectief zou kunnen zijn.
- *De kwaliteit van uitgevoerd onderzoek en de mate waarin onderzoek laat zien dat de interventie leidt tot de realisering van de gestelde doelen.* Interventies die aan deze criteria voldoen heten

'bewezen effectief'. De criteria die hier moeten gelden zijn nog niet vastgesteld. Wat onder meer nog ontbreekt is een goede regel die aangeeft hoeveel effectonderzoeken er op tafel moeten liggen, van welke kwaliteit die moeten zijn en hoe groot het gevonden effect moet zijn om van een 'bewezen effectieve' interventie te mogen spreken.

- *De mate van uitvoerbaarheid en efficiëntie waarmee een interventie zo goedkoop mogelijk, maar met behoud van de effectiviteit in de praktijk is toe te passen (de zogeheten 'kosteneffectiviteit').* Interventies die aan dit criterium voldoen, worden 'doelmatig' of 'kosteneffectief' genoemd. Dit onderdeel van de beoordeling is vooralsnog het minst uitgewerkt. Dit hangt samen met het feit dat veel gegevens die voor de beoordeling op dit punt nodig zijn (zowel met betrekking tot de effectiviteit als de uitvoerbaarheid) op dit moment nauwelijks voorhanden zijn.

De set van criteria is zo beperkt mogelijk gehouden, zodat het geheel overzichtelijk en voor aanbieders begrijpelijk blijft. Daarnaast geldt een aantal kwaliteitscriteria die voor de ontwikkeling van interventies van belang zijn, maar die vooralsnog geen rol spelen bij de erkenning.

De werkwijze is ongeveer hetzelfde als bij die van de hiervoor beschreven databank. Nu zijn het secretarissen van een landelijke erkenningscommissie die over een interventiebeschrijving voorlopig kwaliteitsoordeel geven. Ook in die procedure geeft men naar aanbieders suggesties over de aspecten waarop interventies te verbeteren zijn. De commissie doet vervolgens de definitieve uitspraken. De erkende interventies, alsook de interventies die niet zijn erkend of door de secretarissen in samenspraak met de aanbieder niet aan de commissie zijn voorgelegd worden gepubliceerd op internet, onder andere via www.jeugdinterventies.nl.

De werkgroep heeft nog geen geldigheidsduur van de predikaten vastgesteld; dit gebeurt binnenkort.

3.3 Erkenningscommissie Gedragsinterventies Justitie

De *Erkenningscommissie Gedragsinterventies Justitie* geeft zogeheten 'ex ante' beoordelingen over de effectiviteit van interventies ter voorkoming van recidive van delinquenten. Dit houdt in dat de commissie 'in theorie' nagaat of een interventie ten aanzien van recidivevermindering effectief is te noemen, los van de vraag of deze effectiviteit in de praktijk is bewezen. De commissie hanteert daarbij tien kwaliteitscriteria (zie bijlage 4). Deze hebben onder meer betrekking op de theoretische onderbouwing, de afbakening van de doelgroep, de focus van de interventie, en de aansluiting bij de motivatie van de betrokkene. Tevens geldt als voorwaarde dat er behandelmethoden worden toegepast die bewezen-effectief of veelbelovend zijn en dat er een doorlopende evaluatie moet zijn die inzicht geeft in de effectiviteit van de gedragsinterventie in de praktijk. Er zijn geen expliciete criteria die aangeven wat onder 'bewezen-effectief' moet worden verstaan.

De commissie hanteert drie kwalificaties:

- *Niet erkend.* In die gevallen voldoet de interventie (nog) onvoldoende aan de kwaliteitscriteria.
- *Voorlopig erkende gedragsinterventie.* Een voorlopig erkende gedragsinterventie voldoet nog niet aan alle kwaliteitscriteria. De Erkenningscommissie heeft echter voldoende vertrouwen in de recidiveverminderende werking van de interventie en verwacht dat met enkele noodzakelijke aanpassingen een effectieve interventie ontstaat. De status "Voorlopig erkend" is geldig voor een periode van twee jaar en kan in principe niet worden verlengd.
- *Erkende gedragsinterventie.* Een programma krijgt de status "Erkend" als het aan de tien kwaliteitscriteria voldoet. De status "Erkend" is geldig voor een periode van vijf jaar. Hierna kan het programma opnieuw aan de Erkenningscommissie worden voorgelegd. Dan neemt de commissie ook de gebleken effectiviteit onder de loep.

De werkwijze gaat niet uit van een voorlopige beoordeling met aansluitend een toetsing door de commissie. Erkende interventies worden gepubliceerd via www.justitie.nl; interventies specifiek voor jeugdigen staan ook op www.jeugdinterventies.nl.

3.4 Discussie

Gemeenschappelijk aan de drie trajecten is duidelijk het ontwikkelperspectief: interventies zijn in stappen te kwalificeren in de richting naar de 'evidence-based'-status. De trajecten zoeken daarbij de interactie met de aanbieders, bij de één wat explicieter dan bij de andere.

De criteria zijn momenteel sterk gefocused op de 'ex ante' beoordeling: de nadruk ligt op de vraag of een interventie in principe effectief zou kunnen zijn, gegeven het doel, de doelgroep, de aanpak en de theoretische onderbouwing. De criteria ten aanzien van de 'bewezen effectiviteit' zijn in de trajecten nog niet goed uitgewerkt. In de Databank Effectieve Jeugdinterventies is daar een voorlopige classificatie voor ontwikkeld. Die classificatie is echter beschrijvend van aard en niet normerend: er is bijvoorbeeld niet aangegeven hoeveel effectstudies er moeten zijn van welke kwaliteit en met welke resultaten om van 'bewezen effectief' te mogen spreken.

Wat voorts opvalt is dat in geen van de trajecten bij de erkenning criteria aanwezig zijn ten aanzien van de kosteneffectiviteit. Bij dit aspect gaat het – kort gezegd - om de verhouding tussen wat een interventie kost en wat het oplevert.

4. Andere systemen

Plaatst men de drie genoemde trajecten tegen de achtergrond van literatuur en databanken in het buitenland, dan is duidelijk dat er een forse variëteit bestaat in gehanteerde criteria en predikaten. Bijlagen 5 tot en met 8 geven hiervan een bloemlezing. Die variëteit geldt ook ten aanzien van de criteria die zijn aan te leggen voor het predikaat 'bewezen-effectief'. Zoals eerder opgemerkt kan dat te maken hebben met het feit dat de criteria afkomstig zijn uit uiteenlopende sectoren, met elk eigen tradities, wetenschappelijke maatstaven en standen van ontwikkeling.

De vraag is niettemin of integratie van al deze verschillende systemen mogelijk is. In theorie is een dekkend systeem te krijgen door een optelsom te maken van en – waar bij bepaalde criteria zich varianten voordoen – de grootste gemene deler te halen uit alle voorkomende criteria. De uiteenlopende systemen overziend, doen zich hierbij echter drie vraagstukken voor.

4.1 Uitputtende criteria?

De optelsom van criteria leidt tot een flinke lijst. De ervaring leert bovendien dat bespreking van een dergelijke lijst vaak leidt tot nog verdere detaillering en aanvulling. Het eerste vraagstuk betreft dan ook de mate waarin de criteria uitputtend moeten zijn. Bij de beoordeling van de kwaliteit van interventies en het daaraan gekoppelde effectonderzoek zijn in principe talloze criteria te stellen. Twee voorbeelden illustreren dit:

- Doelen van de interventie moeten zijn gespecificeerd. Aan deze doelen zijn weer kwaliteitscriteria te koppelen, zoals de mate waarin ze specifiek, meetbaar, aanvaardbaar, realistisch en tijdgebonden (SMART) zijn.
- De effectiviteit van een interventie moet zijn onderzocht bij een groep jeugdigen en/of opvoeders. Daarbij zijn weer eisen te stellen ten aanzien van de mate van representativiteit van de onderzoeksgroep voor de doelgroep van de interventie, de minimale omvang van de groep, de maximaal toegestane uitval uit de onderzoeksgroep, etc.

Geen enkele interventie zal voldoen aan alle criteria. Hantering van de optelsom van eisen maakt de erkenning tot het najagen van een ideaal, dat weinig resultaten oplevert. Het vergroot bovendien de distantie tot de aanbieders.

Productiever is het in dat verband – net als bij het traject van de Erkenning Jeugdinterventies - onderscheid te maken in minimaal te stellen criteria en aanvullende kwaliteitseisen. Daar waar relevant worden de criteria en kwaliteitseisen nader gespecificeerd voor de toepassing op specifieke werkvelden.

Wanneer vanuit dit oogpunt naar de uiteenlopende systemen wordt gekeken dan valt op dat het in essentie steeds gaat om dezelfde aandachtspunten: een interventie moet goed zijn beschreven (doel, doelgroep, aanpak), theoretisch goed zijn onderbouwd en het effectonderzoek moet zowel interne als externe validiteit bezitten. De criteria die in de reeds lopende trajecten worden gehanteerd, sluiten bij deze essentie aan. Omwille van de continuïteit ligt het voor de hand om hierop voort te borduren.

4.2 Mogelijkheden van effectonderzoek

Daarnaast is de vraag in welke mate er eisen zijn te stellen aan het type onderzoek en de daaraan verbonden methodologische eisen, gezien de aard van de interventie. Twee voorbeelden ter illustratie:

- Bij een onderzoek naar de medicamenteuze behandeling van kinderen met angststoornissen is het mogelijk gerandomiseerd dubbelblind effectonderzoek te doen. Gegeven die mogelijkheid is de werking van medicijnen bij kinderen op een zeer strenge manier te toetsen.

- Bij een onderzoek naar de effecten van massamediale campagnes die ouders wijzen op hun verantwoordelijkheid om het drinkgedrag van hun kinderen te beperken is een experimentele opzet met een interventie- en een controlegroep lastig te realiseren. Wel is het mogelijk om (met nulmetingen als controleconditie) de kennis en het gedrag van opvoeders en het drinkgedrag van kinderen over langere tijd te meten om na te gaan of dit zich in de voorspelde richting ontwikkelt. Een andere mogelijkheid is een internationale vergelijking te maken tussen soortgelijke landen met en zonder dergelijke campagnes.

De vraag is hier bij welke typen interventies welke eisen zijn te stellen. Voor zover relevant, zal ook hier een onderscheid gemaakt moeten worden in minimale criteria en aanvullende eisen. Minimale criteria zijn reeds in het traject van de Databank Effectieve Jeugdinterventies gespecificeerd. Van belang is in de komende tijd voor verschillende soorten interventies aanvullende eisen te specificeren.

4.3 De gehanteerde normen

Hier gaat het om de vraag wanneer voor een interventie voldoende bewijs is om van ‘bewezen-effectief’ te mogen spreken. Bij het antwoord wegen verschillende zaken mee, zoals:

- De mate waarin men zekerheid wenst over de gevonden effecten. Dit bepaalt bijvoorbeeld in welke mate men buitenlands onderzoek accepteert, van welke kwaliteit het onderzoek moet zijn, hoe betrouwbaar de gebruikte instrumenten moeten zijn, hoeveel effectonderzoeken er moeten zijn, welke effectgrootte die moeten laten zien etc.
- De mate waarin de risico’s van een interventie opwegen tegen het niet-interveniëren of het gebruik van andere interventies. Medicamenteuze behandeling van angst bij kinderen wordt over het algemeen als niet verantwoord beschouwd als ‘slechts’ in theorie aannemelijk is dat deze interventie helpt. Het risico is immers dat er onvoorziene bijwerkingen optreden die op lange termijn schadelijk zijn voor de ontwikkeling van het kind. “Baat het niet dan schaadt het niet’ vormt hier geen regel. De grens voor ‘bewezen effectiviteit’ dient hier wellicht op een strenger niveau te worden gelegd dan bij een faalangstcursus voor middelbare scholieren.
- De balans die men maakt van enerzijds de mate waarin een bepaalde investering in ontwikkeling, onderzoek en implementatie van een interventie nodig is en anderzijds de cliëntgebonden of maatschappelijke effecten deze investering met zich meebrengt.

De vraag is in hoeverre de vaststelling van de normen een exclusieve aangelegenheid is van onderzoekers. Het ligt voor de hand dat wetenschappers de meest aangewezen partij zijn om aan te geven wat er nodig is om een acceptabele graad van zekerheid ten aanzien van de effectiviteit te verkrijgen. Bij het tweede en derde punt ligt dat wat ingewikkelder, omdat hier ook ethische en economische motieven in meewegen.

Daarnaast geldt ook hier de vraag hoeveel interventies straks als ‘bewezen effectief’ te erkennen zijn bij eenmaal vastgestelde normen. Kijken we naar de internationale maatstaven, dan zien we bijvoorbeeld dat het gebruikelijk is om te vereisen dat er minstens twee experimentele onderzoeken gedaan moeten zijn – met een aselecte indeling van subjecten – om te kunnen bepalen of een interventie effectief is of niet. Te verwachten is dat – gegeven de stand van de ontwikkeling van de sector – een dergelijk criterium in Nederland maar weinig erkenningen gaat opleveren. Het lijkt van belang om ook hier in de bulk van (nog) niet bewezen-effectieve programma’s een differentiatie aan te brengen om verdere onderzoeksinvesteringen te kunnen richten.

4.4 Conclusie

De conclusie is dat de integratie van criteria uit moet gaan van een minimaal te hanteren set, met aanvullende kwaliteitseisen. Bij de formulering van deze minimale set is het van groot belang zoveel mogelijk aansluiting te zoeken bij de bestaande erkenningspraktijken in Nederland. Dat moet voorkomen dat er een discontinuïteit ontstaat in reeds ingezette ontwikkelingen.

Aanvullend is het noodzakelijk normen te formuleren om te kunnen vaststellen welke interventies als 'bewezen effectief' te kwalificeren zijn. Daarbij moet rekening gehouden worden met de noodzaak een differentiatie aan te brengen in de gehanteerde normen voor 'bewezen effectiviteit', al naar gelang de specifieke toepassingsgebieden en de stand van ontwikkeling. Deze differentiatie zal in de komende tijd voor een belangrijk deel ontwikkeld moeten worden.

5. Een integraal kader

Op basis van de verschillende systemen is een kader te maken waarin de criteria die te stellen zijn in samenhang zijn gebracht. Uitgangspunt daarbij is dat de erkenning van interventies in een ontwikkelingsperspectief moet worden gezet. Daarin worden vier fasen onderscheiden:

- Ontwikkelingsniveau 0: De interventie voldoet (nog) niet aan basiseisen die eraan te stellen zijn (titel erkenning: Niet erkend).
- Ontwikkelingsniveau I: De interventie voldoet aan basiseisen die eraan te stellen zijn (titel erkenning: In theorie effectief).
- Ontwikkelingsniveau II: Er is overtuigend empirisch bewijs inzake de effectiviteit van de interventie (titel erkenning: Bewezen effectief).
- Ontwikkelingsniveau III: De kosten van de interventie staan in een goede verhouding tot de effecten (titel erkenning: Kosteneffectief).

Het kader schetst vervolgens per ontwikkelingsfase de basiseisen waaraan een interventie minimaal moet voldoen. Daarnaast schetst het belangrijke aanvullende kwaliteitseisen alsook de uiteenlopende kwalificaties die toe te kennen zijn als het gaat om de bewezen effectiviteit. De verschillende onderdelen worden hieronder kort toegelicht. In een afsluitende paragraaf is het een en ander in een samenvattend model bijeengebracht.

5.1 Ontwikkelingsniveau 0

Hier gaat het om de interventies die niet voldoen aan de basiscriteria die aan interventies te stellen zijn. De beschrijving van ontwikkelingsniveau I maakt duidelijk om welke basiscriteria het hier gaat.

5.2 Ontwikkelingsniveau I: in theorie effectief

Bij ontwikkelingsniveau I gaat het om de kwaliteit van de beschrijving, onderbouwing en de uitvoerbaarheid van interventies. De minimale vereisten die hier aan de orde zijn, gelden als *erkenningcriteria*. Daarnaast zijn er *kwaliteitscriteria* waaraan een interventie (vooral nog) niet hoeft te voldoen om een erkenning op dit niveau te verkrijgen. Het zijn belangrijke aspecten van kwaliteit, volgens de geraadpleegde systemen. Ze zijn daarom toch genoemd, met het oog op de mogelijkheid dat ze in een later stadium deel kunnen gaan uitmaken van de erkenningcriteria.

In de onderstaande lijst zijn de criteria kort aangeduid. Daarbij is tevens aangegeven welke speciale aandachtspunten er gelden ten aanzien van het justitiële toepassingsveld.

Beschrijving

Erkenningcriteria

- Alle relevante kenmerken van de doelgroep zijn beschreven.
- Er zijn gegevens over indicatie en contra-indicatie.
Justitie: het type justitiabele waarop de gedragsinterventie zich richt wordt gespecificeerd en geselecteerd.
- Er zijn expliciete doelen geformuleerd, zonodig onderscheiden in voorwaardelijke doelen en einddoelen.
Justitie: de gedragsinterventie is gericht op het veranderen van risicofactoren die samenhangen met het criminele gedrag en de aanpak is mede gericht op het leren van praktische, sociale en probleemoplossende vaardigheden.
- De methodiek/werkwijze is zo volledig mogelijk beschreven op het niveau van concrete activiteiten.
- Volgorde, frequentie, intensiteit, duur en timing van contacten en activiteiten zijn gegeven.
- De benodigde materialen en hun verkrijgbaarheid zijn duidelijk beschreven.

Kwaliteitscriteria

- De doelen zijn specifiek en meetbaar (indien relevant per doelgroepsegment).
- De doelen zijn haalbaar binnen de randvoorwaarden van de interventie.
- De handleiding of het protocol is bruikbaar en werkt ondersteunend bij de uitvoering.
- De ontwikkelaar, licentiehouder, uitvoerende en/of ondersteunende organisatie(s) worden genoemd.
- Gegevens over onderhoud en kwaliteitszorg worden gespecificeerd (bijvoorbeeld: wie is verantwoordelijk, wat wordt gedaan, terugkerend onderhoud).

- Indien de interventie niet in Nederland ontwikkeld is, dan wordt de oorspronkelijke context beschreven en worden de aanpassingen die zijn gemaakt om aan te sluiten bij de Nederlandse situatie toegelicht.

Theoretische onderbouwing

Erkenningscriteria

- De probleemanalyse is voldoende volledig (met gegevens over bijv. aard, ernst, omvang, spreiding, perceptie door betrokkenen, kosten) en omvat een analyse van de met het probleem samenhangende oorzakelijke, risico-, instandhoudende, verzachtende of protectieve factoren.
- Er wordt een sterke theoretische onderbouwing gegeven met een samenhangend betoog over:
 - doelgroepformulering (de aansluiting tussen probleemanalyse en doelgroep), en:
 - doelverantwoording (de aansluiting tussen probleemanalyse en doelen; de te beïnvloeden factoren en determinanten zijn gespecificeerd), en:
 - werkzame factoren (de aansluiting tussen probleemanalyse en methodiek).

Justitie: de interventie is gebaseerd op een expliciet veranderingsmodel waarvan de werking wetenschappelijk is aangetoond.
- Doelgroep, doelen en werkwijze sluiten onderling aan.

Justitie: de intensiteit en duur van de interventie sluit aan bij de problematiek van de deelnemer en betrokkenheid van de deelnemer bij de gedragsinterventie en motivatie voor deelname moeten worden bevorderd en gestimuleerd.
- *Justitie:* er worden (behandel)methoden toegepast die aantoonbaar effectief of veelbelovend zijn.

Kwaliteitscriteria

- Er worden technieken gebruikt die in de literatuur als 'bewezen-effectief' bekend staan. Bijvoorbeeld:
 - ruimte voor persoonlijke benadering;
 - feedback over de eigen situatie, gedrag of effecten;
 - gebruik van beloningsstrategieën;
 - barrières voor gewenst gedrag wegnemen;
 - mobiliseren van sociale steun en betrokkenheid;
 - aanleren van vaardigheden;
 - follow-up regelen;
 - goalsetting en implementatie-intenties;
 - interactieve benadering.
- Waar relevant worden bronvermeldingen gegeven m.b.t. de theoretische onderbouwing.

Randvoorwaarden uitvoering en eisen kwaliteitsbewaking

Erkenningscriteria

- De interventie is overdraagbaar, bijvoorbeeld blijkt een systeem van trainingen, begeleiding, registratie, licenties, etc, een handleiding of protocol voor overdracht / implementatie, eerdere ervaringen t.a.v. de overdraagbaarheid.
- *Justitie:* er moeten duidelijke verbindingen zijn tussen de interventie en de totale begeleiding van de justitiabele.
- *Justitie:* de gedragsinterventie wordt uitgevoerd zoals het bedoeld is.
- *Justitie:* een doorlopende evaluatie geeft inzicht in de effectiviteit van de gedragsinterventie.

Kwaliteitscriteria

- De uitvoerende organisatie (indien bekend) heeft een HKZ-registratie. (Dit is een belangrijk erkenningscriterium, maar kan pas in een later stadium van kracht worden i.v.m. de huidige stand van zaken (voortgang) bij de HKZ-registratie van organisaties.)
- Randvoorwaarden en eisen voor kwaliteitsbewaking van de interventie worden duidelijk beschreven.

Bijvoorbeeld:

 - de kosten van uitvoering, inclusief de kaders waarbinnen deze gelden
 - benodigde personele inzet en functies
 - de benodigde tijdsinvestering
 - opleiding en supervisie uitvoerenden
 - instrumenten voor registratie en evaluatie
 - waarborging van programma-integriteit
 - steun en commitment van samenwerkingspartners en financiers
 - steun en commitment vanuit de organisatie
 - informatie, consultatie en participatie van de organisatie
 - expertise en kenmerken van de projectleider
 - aandachtspunten voor de projectleider
 - omgevingseisen t.a.v. veiligheid, ondersteuning en eventuele aanpassingen
- De uitvoerbaarheid van de interventie kan worden getypeerd en is positief.

Te overwegen is bij dit niveau – net als bij justitie - een onderscheid te maken in 'voorlopig erkend' en 'erkend'. Daarbij is dan expliciet aangegeven aan welke criteria moet zijn voldaan om een voorlopige erkenning te krijgen (bijvoorbeeld: in ieder geval alle criteria ten aanzien van de beschrijving en uitvoering/kwaliteitsbewaking). De status "Voorlopig erkend" is dan geldig voor een periode van twee jaar en kan in principe niet worden verlengd.

5.3 Ontwikkelingsniveau II: bewezen effectief

Ontwikkelingsniveau II heeft betrekking op de empirische onderbouwing van de effectiviteit van de interventie. Bij die empirische onderbouwing spelen twee dimensies een rol: de kwaliteit van de uitgevoerde studie en de overtuigingskracht van het resulterende bewijs.

Kwaliteit van de studie

Voorstel is hiervoor het zogeheten ‘sterrensysteem’ van de Databank Effectieve Jeugdinterventies te gebruiken. Figuur 2 geeft dit systeem kort weer.²

Figuur 2. ‘Sterrensysteem’ kwaliteit van de effectstudie

Sterren	Kenmerken studie	MSMs-niveaus
°	Studies zonder een ster voldoen niet aan de eisen die minstens aan een effectstudie te stellen zijn.	-
*	<ul style="list-style-type: none"> De meting is (mede) gericht op de doelen en de doelgroep van de interventie. De meting is verricht met instrumenten die voldoende betrouwbaar zijn. De meting is verricht met instrumenten die de doelen van de interventie valide operationaliseren. Er is een voormeting (voorafgaand aan / bij start van de interventie). Er is een nameting (aan het einde van de interventie). De resultaten zijn met een adequate statistische techniek geanalyseerd en op significantie getoetst. Er is bovendien eventueel een follow-up (minimaal 6 maanden) en/of de studie is in de praktijk uitgevoerd.	Niveau 1
**	Hier gelden vrijwel dezelfde criteria als bij *, met als verschil: <ul style="list-style-type: none"> Er is een experimentele en een controle-conditie. Er kan sprake zijn van een experimentele en controlegroep, een herhaald N=1 onderzoek met een baseline of een time series design met een baseline. De studie is niet in de praktijk is uitgevoerd. Door de controleconditie is de kans op interne validiteit groter.	Niveau 2-5 (niet in de praktijk)
***	Hier gelden vrijwel dezelfde criteria als bij **, met als verschil: <ul style="list-style-type: none"> De studie is in de praktijk uitgevoerd. De kans op externe validiteit van de gegevens is daardoor groter.	Niveau 2-5 (in de praktijk)
****	Hier gelden vrijwel dezelfde criteria als bij ***, met als verschil: <ul style="list-style-type: none"> Er is een follow-up (minimaal 6 maanden). De kans op externe validiteit en het zicht op de stabiliteit van de gegevens is daardoor groter.	Niveau 2-4 (in de praktijk) (met follow-up)
*****	Hier gelden vrijwel dezelfde criteria als bij ****, met als verschil: <ul style="list-style-type: none"> Er is een a-selectie toewijzing van subjecten aan onderzoeksgroepen. De kans op interne validiteit is daardoor groter.	Niveau 5 (in de praktijk) (met follow-up)

Kenmerk van dit systeem is dat het op een voor aanbieders en beleidsmakers toegankelijke en aansprekende manier helder maakt welke soorten studies er zijn, welke eisen daaraan te stellen zijn en welke functie die verschillende soorten onderzoek hebben bij de ontwikkeling van de effectiviteit van een interventie. Voor een gedetailleerde toelichting zie het zogeheten Bronnenboek Zicht op Effectiviteit.³ Ter illustratie legt het schema in figuur 2 een relatie met de niveaus van de bekende Maryland Scientific Methods scale (MSMs; zie ook bijlage 6). Kenmerk van het sterrensysteem is dat het – in vergelijking met de MSMs - een nadrukkelijker plaats geeft aan de eis dat er uiteindelijk sprake moet zijn van zowel een goede experimentele opzet, alsook van follow-up onderzoek en uitvoering van studies in de praktijk.

Studies met *o sterren* leveren per definitie geen directe indicaties op voor de effectiviteit. Het gaat hier bijvoorbeeld om tevredenheidsonderzoek, onderzoek naar de mate waarin doelgroepen worden bereikt of procesevaluaties. Dit soort studies kunnen aanwijzingen geven voor succes- en faalfactoren bij de uitvoering van de interventie die de effectiviteit kunnen bevorderen of hinderen.

Studies met *één tot drie sterren* worden gezien als onderzoeken die indicatief bewijs kunnen leveren voor de effectiviteit van een interventie. Het gaat hier bijvoorbeeld om veranderingsonderzoek (zonder controlegroep) en quasi-experimenteel onderzoek dat niet in de dagelijkse praktijk is uitgevoerd of

² In de databank heeft het systeem momenteel vier sterren. Dit wordt veranderd conform de opzet die in figuur 3 staat.

³ Yperen, T.A. van & Veerman, J.W. (2006). Bronnenboek Zicht op Effectiviteit. Deel 1. Utrecht/Nijmegen: NIZW Jeugd / Praktikon. Te downloaden van www.nji.nl > zoekterm ‘Bronnenboek’.

waarin geen follow-up is ingebouwd. Over het algemeen geldt dat het verstandig is pas geavanceerder effectonderzoek te doen (bijvoorbeeld van vier of vijf sterren) als er indicatief bewijs is dat laat zien dat verder onderzoek de moeite waard is. Als bewijs van dit niveau al vrij duidelijk laat zien dat de interventie geen of negatieve resultaten lijkt op te leveren, zal dit in de meeste gevallen ertoe leiden dat de interventie wordt aangepast alvorens een geavanceerder onderzoek te starten. Positieve uitkomsten uit dit type onderzoek kunnen voor financiers reden zijn voorrang te geven bij de toekenning van middelen voor nadere studie, ten opzichte van interventies waarover nog in het geheel geen gegevens bekend zijn. Bij dat soort beslissingen kunnen onderzoeken met bijvoorbeeld drie sterren zwaarder worden gewogen dan studies met één ster.

Studies met *vier of vijf sterren* gelden over het algemeen als de 'echte' effectstudies. Het onderscheid in de twee niveaus ligt in de vraag of er geen (vier sterren) of wel (vijf sterren) een a-selectie toewijzing is aan de interventie-of de controle-conditie. Het laatste niveau geldt doorgaans als de beste opzet om causale verbanden te kunnen leggen tussen de inzet van de interventie en de gevonden resultaten.

Erkenningscriteria

Als criterium om in aanmerking te komen voor de kwalificatie voor 'bewezen effectiviteit' geldt ten aanzien van de kwaliteit van de studie:

- Er is sprake van een vier- of vijfsterren onderzoek.

Kwaliteitscriteria

Net als bij het voorgaande niveau gelden ook hier aanvullende kwaliteitseisen. Onderstaande lijst is nog niet volledig. In de toekomst is een selectie van deze eisen mogelijk toe te voegen aan de erkenningseisen.

- Er is ervaring opgedaan met de interventie of er is een pretest of marktverkenning uitgevoerd; hierdoor zijn mogelijke succes- en faalfactoren beschikbaar en/of de interventie is aangepast.
- Er is een procesevaluatie verricht, die inzicht geeft in bijvoorbeeld:
 - de mogelijkheid van uitvoering volgens de beschrijving
 - het oordeel van gebruikers over activiteiten en materialen
 - het bereik van de interventie
 - onvoorziene omstandigheden en neveneffecten
 - succes- en faalfactoren
 - andere aspecten van uitvoeringDe resultaten van deze procesevaluatie waren overwegend positief.
- Er zijn secundaire effecten van de interventie aanwijsbaar. Bijvoorbeeld:
 - procesmatig effecten op doelgroepniveau
 - aangeleerde vaardigheden op doelgroepniveau
 - veranderingen in overheidsbeleid
 - veranderingen in het organisatiebeleid
 - verbetering in intersectorale samenwerking
 - meer inzet en/of middelen op lokaal/regionaal niveau
 - effecten op de omgeving (gezondheid)
- Er is een controlegroep met een andere, duidelijk gespecificeerde interventie.
- Indien er sprake is van een herhaalde N=1-opzet, bestaat het design uit een baseline-controleconditie, een interventiefase, een fase van onthouding van de interventie en een nieuwe interventiefase.
- De mate van uitval van subjecten tussen de meetmomenten is gespecificeerd.
- Het aantal subjecten en metingen resulteren in een adequate statistische power.
- De implementatiegetrouwheid is bepaald (i.e. nagegaan is wat de mate is waarin het protocol, de handleiding of de methodiek getrouw is gevolgd - ook wel behandelingsintegriteit, 'treatment integrity' of 'fidelity' genoemd).
- Het onderzoek is uitgevoerd door anderen dan de ontwikkelaars of de aanbieders van de interventie.

Overtuigingskracht van het bewijs

Voor de criteria ten aanzien van de overtuigingskracht van het bewijs is uit de literatuur op te maken dat hier met name twee aspecten op de voorgrond treden. Het eerste is het aantal studies dat moet uitgevoerd zijn. De APA-normen (zie bijlage 6) hanteren bijvoorbeeld

- voor het predikaat 'Well-Established Treatments': 'At least two good between group design experiments';
- voor 'Possibly efficacious treatments': 'One between group design experiment'.

Het tweede aspect betreft de grootte van het gemeten effect (Effect Size of ES). De meest bekende indeling is die van Cohen's *d*. In theorie kan de index elke waarde aannemen, praktisch

gezien schommelt d meestal tussen de -2.0 en $+2.0$. Onderstaand schema (figuur 3) biedt vuistregels om de waarde van de index te interpreteren.⁴

Figuur 3: Classificatie van de Effect Size (ES) Cohen's d

Een ES tussen	Wijst op
1.3 en hoger	een zeer groot effect
.80 en 1.29	een groot effect
.50 en .79	een middelgroot effect
.20 en .49	een klein effect
-.19 en .19	geen of een verwaarloosbaar effect
-.20 en -.49	een klein negatief effect
etc.	etc.

Bij meer dan één studie is een gemiddelde effectsize te bepalen, waarbij elke effectsize gewogen wordt naar de grootte van de groepen waarop deze is gebaseerd.

Specifiek ten aanzien van de Nederlandse situatie geldt daarbij nog de vraag in welke mate het onderzoek van Nederlandse bodem moet zijn. Immers, buitenlandse effectstudies zijn niet zonder meer representatief voor de toepassing in Nederland.

De criteria ten aanzien van het aantal studies, de effectgrootte en de mate waarin er sprake is van Nederlands onderzoek zijn te combineren tot een classificatieschema voor de overtuigingskracht van het bewijs. Dat schema is te differentiëren naar de kwaliteit van het uitgevoerde onderzoek. Figuur 4 geeft daarvan een voorbeeld.

Figuur 4. Voorbeeld classificatie overtuigingskracht van het bewijs#

Overtuigingskracht bewijs →	Aantal studies	Ze ^e r zwak	matig	redelijk	sterk	Ze ^e r sterk
		(Gewogen gemiddelde) ES				
Kwaliteit onderzoek ↓	*****					
	Studies (> 2) buitenland	< .79	> .80			
	+ 1 NL studie	< .20	.20-.49	.50-.79	.80-1.29	> 1.30
	+ 2 NL studies	< .20	.20-.39	.39-.59	.60-.79	> .80
	+ >2 NL studies	< .10	.10-.29	.30-.49	.50-.69	> .70
	1 NL studie	< .20	.20-.79	> .80		
	2 NL studies	< .20	.20-.49	.50-.79	.80-1.29	> 1.30
	>2 NL studies	< .20	.20-.39	.40-.59	.60-.79	> .80

	Studies (> 2) buitenland	< 1.29	> 1.30			
	+ 1 NL studie	< .40	.40-.99	> 1.00-1.29	> 1.30	
	+ 2 NL studies	< .40	.40-.59	.60-.99	1.00-1.29	> 1.30
	+ >2 NL studies	< .40	.40-49	.50-79	.80-1.00	> 1.00
	1 NL studie	< .50	.50-1.29	> 1.30		
	2 NL studies	< .20	.20-79	.80-1.29	> 1.30	
	>2 NL studies	< .20	.20-59	.60-79	.80-1.29	>1.30

De gegeven ES-waarden zijn slechts bedoeld als illustratie en pretenderen niet exact te zijn.

⁴ Afgeleid van: Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, p.155-159.

Op basis daarvan zijn vervolgens erkenningscriteria en aanvullende te formuleren, bijvoorbeeld:

Erkenningscriteria

- De overtuigingskracht van het bewijs is minstens van redelijk niveau.

Kwaliteitscriteria

- Er zijn minstens twee studies van Nederlandse bodem.

Wellicht ten overvloede moet gezegd worden dat de classificaties en criteria die in figuur 4 zijn aangegeven nog ter discussie moeten staan. De specificaties die hier zijn gegeven dienen als illustratie voor de te ontwikkelen systematiek.

Ook hier is te overwegen een onderscheid te maken in ‘voorlopig erkend’ en ‘erkend’. Bijvoorbeeld, om voorlopig erkend te worden kan als eis gelden: ‘de overtuigingskracht van het bewijs is minstens van matig niveau’.

5.4 Ontwikkelingsniveau III: kosteneffectief

Bij de kosteneffectiviteit van interventies gaat het om de verhouding tussen enerzijds de kosten voor de uitvoering van de aanpak en anderzijds de effecten. Voor de definitie, afbakening van andere economische aspecten van het al dan niet interveniëren en te hanteren rekenmodellen zijn goede bronnen beschikbaar.⁵

Het onderwerp – en daarmee ook het betreffende ontwikkelingsniveau – valt op dit moment buiten het bereik van de erkenning. De reden daarvoor is tweeledig.

- In de eerste plaats veronderstelt deze benadering dat er gegevens zijn over de kosten van interventies (alsook van het niet-interveniëren). De berekening van die kosten staat in een groot deel van de sector echter in de kinderschoenen.⁶ Betrouwbare en trefzekere schattingen zijn onvoldoende voorhanden.
- In de tweede plaats zijn voor berekeningen van de kosteneffectiviteit goede schattingen nodig van de effectiviteit van de interventies. Ook daarover ontbreken in veel gevallen goede gegevens.

Ondanks deze problemen is het van belang het ontwikkelingsstadium waarin de kosteneffectiviteit centraal staat in het vizier te houden. Het vergt dat ook nu al – ter voorbereiding - aandacht wordt besteed aan het op orde brengen van de noodzakelijke gegevens. Dat die gegevens momenteel buiten het bereik van de erkenning vallen, betekent niet dat andere partijen hier met het oog op de toekomst huiswerk in hebben te doen.

5.5 Samenvattend model

Figuur 5 vat het hierboven geschetste kader in een model samen. De verschillende niveaus ‘stapelen’ zich daarin op. Vanaf niveau I gelden er erkenningscriteria die bepalen of een interventie ‘In theorie effectief’ (I), ‘Bewezen effectief’ (II) of ‘Kosteneffectief’ (III) is te noemen. Niveau I bepaalt of een interventie goed beschreven, onderbouwd en uitvoerbaar is. Bij niveau II gaat het om de vraag of een interventie als ‘bewezen effectief’ mag gelden. Voor dit predikaat speelt feitelijk alleen het zogeheten vier- en vijfsterren-onderzoek een rol in de erkenning. Dat maakt ander onderzoek niet overbodig,

⁵ Zie o.a. Scott, S. (2001). Deciding whether interventions for antisocial behaviour work: Principles of outcome assessment, and practice in an multicentre study. *European Child & Adolescent Psychiatry*, 10, 1/59-1/70; Welsh, B. (2001). Kosten en baten van preventie en behandeling van ernstige en gewelddadige jeugdcriminaliteit. In: R. Loeber, N.W. Slot & J.A. Sergeant (Red.), *Ernstige en gewelddadige jeugddelinquentie. Omvang, oorzaken en interventies* (p. 373-390). Houten: Bohn, Stafleu Van Loghum; Wiersma, D., F.G. Brook & S. Sytma (1999). Onderzoek naar het functioneren van de geestelijke gezondheidszorg. In: A. de Jong, W. van den Brink, J. Ormel en D. Wiersma (Red.). *Handboek psychiatrische epidemiologie* (p. 422-441). Maarssen: Elsevier/De Tijdstroom.

⁶ Zie o.a. Yperen, T.A. van, L. Boendermaker en M. Hellema (2004). *Indicatie, modules en bekostiging*. Utrecht: NIZW Jeugd.

maar het levert niet het overtuigende bewijs dat voor de erkenning noodzakelijk is. Bij niveau III staat de kosteneffectiviteit van interventies centraal. Dit aspect is nog onderontwikkeld. Het zal in de toekomst uitgebouwd moeten worden zodat ook hiervoor erkenningen zijn af te geven.

Figuur 5. Integraal kader erkenning interventies

Kosteneffectief: de interventie is minstens redelijk kosteneffectief						
Ontwikkelings-niveau III	<ul style="list-style-type: none"> • Hoge kosteneffectiviteit • Redelijke kosteneffectiviteit • Lage kosteneffectiviteit 					
Bewezen effectief: er is beslissend bewijs met minstens redelijke overtuigingskracht						
Ontwikkelings-niveau II	Overtuigingskracht bewijs →	<i>Zeer zwak</i>	<i>matig</i>	<i>redelijk</i>	<i>sterk</i>	<i>Zeer sterk</i>
	↓ Kwaliteit onderzoek					
	*****	(Beslissend bewijs / 'conclusive evidence')				

	**	(Indicatief bewijs / 'preliminary evidence')				
*						
o	(Geen bewijskracht t.a.v. effectiviteit)					
In theorie effectief: de interventie voldoet aan de basale criteria						
Ontwikkelings-niveau I	Basale criteria voor interventies: <ul style="list-style-type: none"> • Beschrijving • Theoretische onderbouwing • Uitvoerbaarheid en kwaliteitsbewaking 					
Niet erkend: voldoet niet aan criteria niveau I						
Ontwikkelings-niveau 0	Zie criteria niveau I					

6. Afspraken Erkenning Jeugdinterventies voor 2007

De commissie voor de erkenning van jeugdinterventies heeft in het licht van het bovenstaande op 20 juni 2007 de volgende afspraken gemaakt:

- Voor de erkenning van interventies op ontwikkelingsniveau I (*In theorie effectief*) gelden de criteria zoals beschreven in paragraaf 5.2.
- Voor de erkenning van interventies op ontwikkelingsniveau II (*Bewezen effectief*) geldt:
 - Voor de aanduiding van de kwaliteit van het onderzoek hanteert de commissie het 'sterrenstelsel' zoals beschreven in paragraaf 5.3.
 - Voor de aanduiding van de overtuigingskracht van het bewijs onderzoekt de commissie in 2007 welke invulling het kader zoals geschetst in figuur 4 in paragraaf 5.3 concreet moet krijgen (eventueel per werkveld of toepassing). Dit resulteert in een set van voorlopige criteria waarmee de commissie ervaring op doet voor de erkenning op niveau II.
 - Tot 2008 beoordeelt de commissie uitsluitend of interventies voor een erkenning op niveau I in aanmerking komen. Binnen twee jaar, uiterlijk 2009, worden alle in 2007 beoordeelde interventies versneld opnieuw beoordeeld aan de hand van inmiddels ontwikkelde criteria voor niveau II.
- De commissie geeft voorlopig geen erkenningen af voor interventies op ontwikkelingsniveau III (*Kosteneffectief*) om de reden zoals geschetst in paragraaf 5.4.

Bijlagen

Bijlage 1. Criteria Databank effectieve Jeugdinterventies

(Stand van zaken mei 2007)

De criteria van de DEI zijn ontstaan door harmonisatie van gehanteerde criteria voor de beoordeling van programma's voor Voor- en Vroegschoolse Educatie (VVE), integrale aanpak van problemen in de wijk (Communities that Care) en interventies in de jeugdzorg.

Voor uitgebreide informatie, zie www.jeugdinterventies.nl

De Databank toetst in eerste instantie of het gaat om een jeugdinterventie. Zo ja, dan gaat de toetsing verder. Zo nee, dan stopt de toetsing. De volgende criteria spelen daarbij een rol.

- Bevat de documentatie informatie over het doel van de interventie?
- Bevat de documentatie informatie over de doelgroep?
- Bevat de documentatie informatie over de methodiek?
- Bevat de documentatie informatie over het doel van de interventie?
- Is de doelgroep jeugdigen van 0-18 jaar, hun opvoeder of opvoedomgeving?
- Betreft het een interventie (zoals in de handleiding gedefinieerd)?

Daarna wordt er aan de hand van de volgende criteria getoetst op de vraag of de interventie als 'Veelbelovend' is aan te merken.

- Zijn er expliciet doelen geformuleerd?
- Bevat de documentatie een definitie van de doelgroep met demografische kenmerken en/of aard van de problematiek?
- Bevat de methodiek een protocol of handleiding dat de benodigde handelingen (volgorde, duur, frequentie, intensiteit) en materialen aangeeft?
- Is het probleem omschreven?
- Is duidelijk wie het probleem heeft?
- Is een analyse gemaakt van de met het probleem samenhangende oorzakelijke, risico-, in stand houdende, verzachtende of protectieve factoren?
- Is de verantwoording gebaseerd op een samenhangend betoog (een theorie) over met het probleem samenhangende factoren en de veronderstelde werkzame ingrediënten in de interventie (wordt er méér beschreven dan een algemene verwijzing naar een theorie zoals "De interventie is gebaseerd op de sociale leertheorie")?
- Sluit de keuze van de doelen van de interventie aan bij de probleemanalyses?
- Wordt aannemelijk gemaakt dat de veronderstelde werkzame factoren in de methodiek leiden tot het bereiken van de genoemde doelen (blijkt uit onderzoek, theorie en / of praktijk dat de methodiek de beoogde doelen bij de doelgroep kan bereiken)?

De databank toetst niet op effectiviteit, maar geeft een beschrijving van het type uitgevoerde onderzoek en de resultaten daarvan. Beschrijving van het type uitgevoerde onderzoek gebeurt aan de hand van een sterrenstelsel.

Sterren	Kenmerken studie
°	Studies zonder een ster voldoen niet aan de eisen die minstens aan een effectstudie te stellen zijn.
*	<ul style="list-style-type: none"> • De meting is (mede) gericht op de doelen en de doelgroep van de interventie. • De meting is verricht met instrumenten die voldoende betrouwbaar zijn. • De meting is verricht met instrumenten die de doelen van de interventie valide operationaliseren. • Er is een voormeting (voorafgaand aan / bij start van de interventie). • Er is een nameting (aan het einde van de interventie). • De resultaten zijn met een adequate statistische techniek geanalyseerd en op significantie getoetst. Er is bovendien eventueel een follow-up (minimaal 6 maanden) en /of de studie is in de praktijk uitgevoerd.
**	Als bij vorige niveau, behalve dat de studie niet in de praktijk is uitgevoerd. Bovendien geldt: <ul style="list-style-type: none"> • Er is een experimentele en een controle-conditie. Deze opzet van de experimentele en controle-conditie kan van uiteenlopende kwaliteit zijn.
***	Als bij vorige niveau, behalve dat er geen follow-up is. Bovendien geldt: <ul style="list-style-type: none"> • De studie is in de praktijk uitgevoerd. De kans op externe validiteit van de gegevens is daardoor groter.
****	Als het vorige niveau, maar nu geldt ook: <ul style="list-style-type: none"> • Er is een follow-up (minimaal 6 maanden). De kans op externe validiteit en het zicht op de stabiliteit van de gegevens is daardoor groter.
*****	Als het vorige niveau, maar nu geldt ook: <ul style="list-style-type: none"> • Er is een a-selectie toewijzing van subjecten aan de onderzoeksgroepen. NB dit niveau is nog niet in de website van de databank doorgevoerd. Dat gebeurt in de loop van 2007.

Beschrijving van de resultaten per studie gebeurt met onderstaande rubricering (de zogeheten 'Tabel C').

- 0 Geen van de onderstaande rubrieken zijn van toepassing (licht toe!).
- 1 *Effectief*: De studie rapporteert positieve effecten[#] ten aanzien van alle doelen van de interventie.
- 2 *Deels effectief*: De studie rapporteert positieve effecten[#] ten aanzien van een aantal doelen, bij andere doelen geen effecten.
- 3 *Effectiviteit niet aangetoond*: De studie rapporteert dat er geen effecten[#] ten aanzien van de doelen van de interventie zijn.
- 4 *Negatief effect*: De studie rapporteert negatieve effecten[#].
- 8 *Effectiviteit onduidelijk of onbekend*.

[#] Positief effect = een doel van de interventie wordt (deels) gerealiseerd en deze winst is statistisch significant. Geen effect = een doel van de interventie wordt niet of deels gerealiseerd en deze winst is niet statistisch significant. Negatief effect: de interventie werkt - statistisch significant - averechts of heeft ernstige, duidelijk aantoonbare 'bijwerkingen'.

De resultaten uit de verschillende onderzoeken worden vervolgens gekenmerkt aan de hand van een voorlopig classificatiesysteem. Daarbij wordt onderscheid gemaakt in direct bewijs (Nederlands onderzoek naar de betreffende interventie) en indirect bewijs (onderzoek naar soortgelijke interventies of buitenlands onderzoek).

A. Directe aanwijzingen voor de effectiviteit

- 0. Er is *geen Nederlands onderzoek* naar de effectiviteit van de hier beschreven interventie.
- 1. Er is *minstens één Nederlands onderzoek (geen veranderings- of (quasi-)experimenteel onderzoek^o)* dat voor de hier beschreven interventie positief effect^A laat zien.
- 2. Er is *minstens één Nederlands veranderingsonderzoek** dat voor de hier beschreven interventie positief effect^A laat zien.
- 3. Er is *minstens één Nederlands (quasi-) experimenteel onderzoek*** dat voor de hier beschreven interventie positief effect^A laat zien.
- 4. Er is *minstens één Nederlands (quasi-) experimenteel onderzoek in de praktijk**** dat voor de hier beschreven interventie positief effect^A laat zien.
- 5. Er is *minstens één Nederlands (quasi-) experimenteel onderzoek in de praktijk en met follow-up***** dat voor de hier beschreven interventie positief effect^A laat zien.
- 8. Overig. De typering past niet bij rubriek 0-5 (licht toe). Gebruik deze bijvoorbeeld als de makers of uitvoerders bepaalde studies aanvoeren, maar andere belangrijke studies niet, of als er veel onduidelijkheden zijn.
- 9. Het is (vooralsnog) onbekend of er directe aanwijzingen voor de effectiviteit zijn, vanwege problemen met de beschikbaarheid of de interpretatie van bronnenmateriaal.

B. Indirecte aanwijzingen voor de effectiviteit

- 0. Er is *geen onderzoek* aangevoerd dat de theorie achter de aanpak empirisch onderbouwt. Er zijn *geen effectstudies* naar soortgelijke interventies aangevoerd.
- 1. Er is onderzoek aangevoerd dat de theorie achter de aanpak van de interventie empirisch onderbouwt. Of er is *onderzoek van onbekende aard* naar soortgelijke interventies aangevoerd, dat een positief effect^B laat zien. Er zijn verder *geen veranderingsonderzoeken of (quasi-) experimentele effectstudie* naar soortgelijke interventies bekend.
- 2. Er is *veranderingsonderzoek en/of hoogstens één (quasi-) experimentele effectstudie, al dan niet in de praktijk of met follow-up*, dat een positief effect^B laat zien.
- 3. Er zijn *twee of meer (quasi-) experimentele effectstudies, al dan niet met follow-up*, naar soortgelijke interventies uitgevoerd, die een positief effect^B laten zien.
- 4. Er zijn *twee of meer (quasi-) experimentele effectstudies met follow-up* naar soortgelijke interventies uitgevoerd, die een positief effect^B laten zien.
- 5. Er zijn *vier of meer (quasi-) experimentele effectstudies met follow-up* naar soortgelijke interventies uitgevoerd, die een positief effect^B laten zien.
- 8. Overig. Omschrijf kort de aard en de resultaten van de aangevoerde studies.
- 9. Het is (vooralsnog) onbekend of er indirecte aanwijzingen voor de effectiviteit zijn, vanwege problemen met de beschikbaarheid of de interpretatie van bronnenmateriaal.

^A **Positief effect** = Voor de studie(s) is in Tabel C optie 1 of 2 aangekruist *en* er zijn naar schatting geen of verhoudingsgewijs weinig (< 20% van het totale aantal) studies die geen of negatief effect uitwijzen. Indien er studies zijn die geen of negatief effect laten zien en de gemiddelde effectgrootte *d* is bekend, dan is deze groter dan .20.

^B **Positief effect** = Door de interventie zijn de cliënten beter af (indien gemiddelde effectgrootte bekend: $d > .20$). Indien gemiddelde *d* niet bekend: er zijn naar schatting geen of verhoudingsgewijs weinig (< 20% van het totale aantal) studies die geen of negatief effect uitwijzen.

Bijlage 2. Criteria Preffi⁷

Preffi staat voor Preventie Effectmanagement Instrument. Het is een kwaliteitsinstrument dat een systematische werkwijze ondersteunt en tot doel heeft de effectiviteit van gezondheidsbevorderende projecten te vergroten.

Preffi 2.0 is door het NIGZ ontwikkeld in samenwerking met GVO-praktijkmedewerkers, GVO- en preventiedeskundigen en wetenschappelijk onderzoekers.

Preffi 2.0 bestaat uit 8 clusters uit zowel de theorie als de praktijk:

Ontwikkelingsgerichte aspecten

2. Probleemanalyse
3. Determinanten
4. Doelgroep
5. Doelen
6. Interventie-ontwikkeling

Uitvoeringsgerichte aspecten

7. Implementatie
8. Evaluatie
1. Randvoorwaarden en haalbaarheid

Ieder cluster bevat een aantal criteria waarop een project kan scoren met zwak, matig of sterk. Met behulp van deze scores krijgen alle clusters en het totale project een rapportcijfer. Op basis van deze beoordeling kunnen verbeterpunten en acties worden geformuleerd. De criteria hebben betrekking op het volgende:

Analyse

2 Probleemanalyse

- 2.1 Aard, ernst en omvang van het probleem
- 2.2 Spreiding van het probleem
- 2.3 Hoe zien de verschillende betrokkenen het probleem?

3 Determinanten van (psychische) problematiek, gedrag en omgeving

- 3.1 Theoretisch model
- 3.2 Bijdrage van determinanten aan problematiek, gedrag of omgevingsfactor
- 3.3 Beïnvloedbaarheid van de determinant
- 3.4 Prioritering en keuze

Interventiekeuze en -ontwikkeling

4 Doelgroep

- 4.1 Algemene en demografische kenmerken van de doelgroep
- 4.2 Motivatie en mogelijkheden van de doelgroep
- 4.3 Bereikbaarheid van de doelgroep

5 Doelen

- 5.1 Doelen sluiten aan op de analyse
- 5.2 Doelen zijn specifiek, tijdgebonden en meetbaar
- 5.3 Doelen zijn aanvaardbaar
- 5.4 Doelen zijn haalbaar

6 Interventieontwikkeling

- 6.1 Onderbouwing van de rationale van de interventiestrategie
 - 6.1a Afstemming van de strategieën en methoden op doelen en doelgroepen
 - 6.1b Eerdere ervaringen met de interventiemethode
- 6.2 Duur, intensiteit en timing
 - 6.2a Duur en intensiteit van de interventie
 - 6.2b Timing van de interventie
- 6.3 Afstemming op de doelgroep
 - 6.3a Participatie van de doelgroep
 - 6.3b Afstemming op de 'cultuur'

⁷ Bron: Molleman, G., Peters, L., Hommels, L., Ploeg, M. (2003). Preffi 2.0. Operationalisering en normering. Woerden: NIGZ.

- 6.4 Effectieve technieken (aanbevolen)
 - Ruimte voor persoonlijke benadering
 - Feedback over effecten
 - Gebruik van beloningsstrategieën
 - Barrières voor gewenst gedrag wegnemen
 - Sociale steun regelen, omgeving erbij betrekken
 - Aanleren van vaardigheden
 - Follow-up regelen
 - Goal-setting en implementatie-intenties
 - Interactieve benadering
- 6.5 Haalbaarheid in de praktijk
 - 6.5a Afstemming op intermediaire doelgroepen
 - 6.5b Kenmerken van de implementeerbaarheid van de interventies
- 6.6 Samenhang
- 6.7 Pretest

7 Implementatie

- 7.1 Keuze voor implementatiestrategie gericht op intermediairen
 - 7.1a Wijze van implementeren: top down en/of bottom up
 - 7.1b Afstemmen van implementatie-interventies op intermediairen
 - 7.1c Geschiktheid van de aanbieder voor intermediairen
- 7.2 Monitoren en genereren van feedback
- 7.3 Inbedden in een bestaande structuur

Evaluatie

8 Evaluatie

- 8.1 Duidelijkheid en overeenstemming over de uitgangspunten van de evaluatie
- 8.2 Procevaluatie
- 8.3 Effectevaluatie
 - 8.3a Is (of wordt) er een verandering gemeten?
 - 8.3b Is aannemelijk dat de verandering is teweeggebracht door de interventie?
- 8.4 Feedback aan betrokkenen

Randvoorwaarden en haalbaarheid

1 Randvoorwaarden en haalbaarheid

- 1.1 Draagvlak
- 1.2 Capaciteit
- 1.3 Sturing door de projectleider
 - 1.3a Expertise en eigenschappen van de projectleider
 - 1.3b Aandachtspunten voor sturing

Bijlage 3. Criteria Werkgroep Erkenning Jeugdinterventies

De *Werkgroep Erkenning Jeugdinterventies* heeft op basis van met name de criteria van de Databank effectieve Jeugdinterventies en Preffi de volgende set van criteria gedifferentieerd.

Inclusiecriteria

Voorafgaand aan de beoordeling door de erkenningscommissie worden de inclusiecriteria gebruikt voor een snelle 'check' of een interventie in aanmerking komt voor beoordeling. Aan elk van de inclusiecriteria moet voldaan zijn. De contactpersonen voeren deze controle uit, voordat de interventie aan de commissie wordt voorgelegd.

- Er is voldoende documentatie.
Dit betekent:
 - een beschrijving van de interventie waarin doelgroep en doel gegeven zijn
 - een Nederlandse handleiding of protocol
 - Nederlandse gegevens over onderzoek, evaluatie of ervaring
- Er is een persoon of organisatie die informatie kan geven.
- Voldoet aan de definitie van interventie (doelgericht, planmatig en toegevoegd).
- Doel is risico's of problemen te verminderen of te voorkomen, of de ontwikkeling te stimuleren.
- Doelgroep is kinderen en jeugdigen van -9 maanden tot 18 jaar (uitloop tot 24 jaar) en/of hun opvoeders of opvoedingsomgeving.

De werkgroep heeft twee typen criteria geformuleerd: erkenningscriteria en kwaliteitscriteria. De laatstgenoemde zijn criteria waaraan een interventie (vooralsnog) niet hoeft te voldoen om een erkenning te kunnen verkrijgen. Het zijn echter wel belangrijke aspecten van kwaliteit van een interventie, volgens de geraadpleegde kwaliteitssystemen. Ze zijn daarom toch opgenomen in de plannen, met het oog op de mogelijkheid dat ze in een later stadium deel kunnen gaan uitmaken van de erkenningscriteria.

Volledigheid van de beschrijving en degelijkheid van de methodiek

Beschrijving

Erkenningscriteria

- Alle relevante kenmerken van de doelgroep zijn beschreven.
- Er zijn gegevens over indicatie en contra-indicatie.
- Er zijn expliciete doelen geformuleerd, zonedig onderscheiden in voorwaardelijke doelen en einddoelen.
- De methodiek/werkwijze is zo volledig mogelijk beschreven op het niveau van concrete activiteiten.
- Volgorde, frequentie, intensiteit, duur en timing van contacten en activiteiten zijn gegeven.
- De benodigde materialen en hun verkrijgbaarheid zijn duidelijk beschreven.

Kwaliteitscriteria

- De doelen zijn specifiek en meetbaar (indien relevant per doelgroepsegment).
- De doelen zijn haalbaar binnen de randvoorwaarden van de interventie.
- De handleiding of het protocol is bruikbaar en werkt ondersteunend bij de uitvoering.
- Er worden technieken gebruikt die in de literatuur als 'bewezen-effectief' bekend staan. Bijvoorbeeld:
 - ruimte voor persoonlijke benadering;
 - feedback over de eigen situatie, gedrag of effecten;
 - gebruik van beloningsstrategieën;
 - barrières voor gewenst gedrag wegnemen;
 - mobiliseren van sociale steun en betrokkenheid;
 - aanleren van vaardigheden;
 - follow-up regelen;
 - goalsetting en implementatie-intenties;
 - interactieve benadering.
- De ontwikkelaar, licentiehouders, uitvoerende en/of ondersteunende organisatie(s) worden genoemd.
- Gegevens over onderhoud en kwaliteitszorg worden gespecificeerd (bijvoorbeeld: wie is verantwoordelijk, wat wordt gedaan, terugkerend onderhoud).
- Indien de interventie niet in Nederland ontwikkeld is, dan wordt de oorspronkelijke context beschreven en worden de aanpassingen die zijn gemaakt om aan te sluiten bij de Nederlandse situatie toegelicht.

Theoretische onderbouwing

Erkenningscriteria

- De probleemanalyse is voldoende volledig (met gegevens over bijv. aard, ernst, omvang, spreiding, perceptie door betrokkenen, kosten) en omvat een analyse van de met het probleem samenhangende oorzakelijke, risico-, instandhoudende, verzachtende of protectieve factoren.
- Er wordt een sterke theoretische onderbouwing gegeven met een samenhangend betoog over:
 - doelgroepformulering (de aansluiting tussen probleemanalyse en doelgroep), en:
 - doelverantwoording (de aansluiting tussen probleemanalyse en doelen; de te beïnvloeden factoren en determinanten zijn gespecificeerd), en:
 - werkzame factoren (de aansluiting tussen probleemanalyse en methodiek).
- Doelgroep, doelen en werkwijze sluiten onderling aan.

Kwaliteitscriteria

- Waar relevant worden bronvermeldingen gegeven m.b.t. de theoretische onderbouwing.

Randvoorwaarden uitvoering en eisen kwaliteitsbewaking

Erkenningscriteria

- De uitvoerende organisatie (indien bekend) heeft een HKZ-registratie. (Dit criterium gaat pas gelden in een later stadium, i.v.m. de huidige stand van zaken (voortgang) bij de HKZ-registratie van organisaties.)
- De interventie is overdraagbaar, bijvoorbeeld blijkt een systeem van trainingen, begeleiding, registratie, licenties, etc, een handleiding of protocol voor overdracht / implementatie, eerdere ervaringen t.a.v. de overdraagbaarheid.

Kwaliteitscriteria

- Randvoorwaarden en eisen voor kwaliteitsbewaking van de interventie worden duidelijk beschreven. Bijvoorbeeld:
 - de kosten van uitvoering, inclusief de kaders waarbinnen deze gelden
 - benodigde personele inzet en functies
 - de benodigde tijdsinvestering
 - opleiding en supervisie uitvoerenden
 - instrumenten voor registratie en evaluatie
 - waarborging van programma-integriteit
 - steun en commitment van samenwerkingspartners en financiers
 - steun en commitment vanuit de organisatie
 - informatie, consultatie en participatie van de organisatie
 - expertise en kenmerken van de projectleider
 - aandachtspunten voor de projectleider
 - omgevingsisen t.a.v. veiligheid, ondersteuning en eventuele aanpassingen

Doelmatigheid

Erkenningscriteria

- De doelmatigheid van de interventie kan worden getypeerd en is positief. (Het criterium van doelmatigheid wordt in een later stadium meegenomen in de beoordelingen (voorlopige planning; vanaf 2011) omdat hierover voor interventies in Nederland nog nauwelijks gegevens zijn.)

Criteria voor ‘Bewezen effectief’

Erkenningscriteria

Direct bewijs effectiviteit

- Er is Nederlands onderzoek beschreven naar de effectiviteit van de bedoelde interventie (inclusief kenmerken en resultaten).
- Het effect van de interventie is gegeven in getallen en is positief.
- De mate van effectiviteit van de interventie is weergegeven (bijvoorbeeld in de vorm van een *effectsize*).
- Het aantal onderzoeken dat wijst op positief/afwezig/negatief effect kan gespecificeerd worden; dit saldo is positief.

Kwaliteitscriteria

Indirect bewijs effectiviteit

- Er is buitenlands onderzoek beschreven (incl. kenmerken en resultaten) naar de effectiviteit of naar de theoretische onderbouwing van de interventie en de resultaten daarvan waren overwegend positief.
- Er is Nederlands onderzoek beschreven (incl. kenmerken en resultaten) naar soortgelijke interventies en de resultaten daarvan waren overwegend positief.

Ervaringsgegevens

- Er is al ervaring opgedaan met de interventie of er is een pretest of marktverkenning uitgevoerd; hierdoor zijn mogelijke succes- en faalfactoren beschikbaar en/of de interventie is aangepast.

- Er is een procesevaluatie verricht, die inzicht geeft in bijvoorbeeld:
 - de mogelijkheid van uitvoering volgens de beschrijving
 - het oordeel van gebruikers over activiteiten en materialen
 - het bereik van de interventie
 - onvoorziene omstandigheden en neveneffecten
 - succes- en faalfactoren
 - andere aspecten van uitvoering
 De resultaten van deze procesevaluatie waren overwegend positief.
- Er zijn secundaire effecten van de interventie aanwijsbaar.

Bijvoorbeeld:

 - procesmatig effecten op doelgroepniveau
 - aangeleerde vaardigheden op doelgroepniveau
 - veranderingen in overheidsbeleid
 - veranderingen in het organisatiebeleid
 - verbetering in intersectorale samenwerking
 - meer inzet en/of middelen op lokaal/regionaal niveau
 - effecten op de omgeving (gezondheid)

Bijlage 4. Criteria Erkenningscommissie Gedragsinterventies Justitie⁸

Kwaliteitscriteria voor de (ex-ante) beoordeling van gedragsinterventies

1. **Theoretische onderbouwing:** de gedragsinterventie is gebaseerd op een expliciet veranderingsmodel waarvan de werking wetenschappelijk is aangetoond.
2. **Selectie van justitiabelen:** het type justitiabele waarop de gedragsinterventie zich richt wordt duidelijk gespecificeerd en geselecteerd.
3. **Dynamische criminogene factoren:** de gedragsinterventie is gericht op het veranderen van risicofactoren die samenhangen met het criminele gedrag.
4. **Effectieve (behandel)methoden:** er worden (behandel)methoden toegepast die aantoonbaar effectief of veelbelovend zijn.
5. **Vaardigheden en protectieve factoren:** de aanpak is mede gericht op het leren van praktische, sociale en probleemoplossende vaardigheden.
6. **Fasering, intensiteit en duur:** de intensiteit en duur van de gedragsinterventie sluit aan bij de problematiek van de deelnemer.
7. **Betrokkenheid en motivatie:** betrokkenheid van de deelnemer bij de gedragsinterventie en motivatie voor deelname moeten worden bevorderd en gestimuleerd.
8. **Continuïteit:** er moeten duidelijke verbindingen zijn tussen de gedragsinterventie en de totale begeleiding van de justitiabele.
9. **Interventie-integriteit:** de gedragsinterventie wordt uitgevoerd zoals het bedoeld is.
10. **Evaluatie:** een doorlopende evaluatie geeft inzicht in de effectiviteit van de gedragsinterventie.

1. Theoretische onderbouwing

De (strafrechtelijke) gedragsinterventie is gebaseerd op een expliciet veranderingsmodel, waarvan de werking wetenschappelijk is aangetoond. Het moet duidelijk zijn op grond van welke theorie de gedragsinterventie relevante veranderingen tot stand brengt in het gedrag van de delinquenten. De theorie legt een duidelijke relatie tussen delinquent gedrag en het mechanisme waardoor dit gedrag ontstaat, in stand wordt gehouden of wordt versterkt. De werkzame bestanddelen van de gedragsinterventie zijn gerelateerd aan deze mechanismen.

Het gaat dus om het theoretisch fundament van de strafrechtelijke gedragsinterventie, waarbij wordt aangegeven hoe het proces dat tot delinquentie leidt verloopt en hoe men door middel van de gedragsinterventie dit proces denkt te veranderen teneinde hiermee recidive te voorkomen.

Op basis daarvan dient de inhoud van de gedragsinterventie beschreven te zijn.

Duidelijk wordt:

- voor wie de gedragsinterventie is bedoeld (omschrijving probleemgedrag waarop de gedragsinterventie is gericht, risicofactoren die hiermee samenhangen, algemene en demografische kenmerken van de deelnemers zoals leeftijd, sekse, geslacht, etniciteit e.d.)
- welke dynamische risicofactoren de gedragsinterventie zal verminderen (doelstelling)
- op welke wijze de doelstelling (vermindering van de invloed van de risicofactoren) wordt bereikt
- welke (behandel)methoden worden toegepast
- wat er na elke fase in de gedragsinterventie wordt bereikt (concrete doelen)
- waarom de doelstelling wordt bereikt door juist deze combinatie van doelen en methoden voor deze doelgroep.

Wanneer een theoretisch concept of wetenschappelijk bewijs voor de werking van een (strafrechtelijke) gedragsinterventie onvolledig is of vooralsnog ontbreekt (bijvoorbeeld bij nieuw ontwikkelde gedragsinterventies), dan wordt het veranderingsmodel aangegeven in de vorm van waarschijnlijke hypothesen.

Duidelijk wordt:

- door welke wetenschappelijke bevindingen de aanpak wordt ondersteund.

2. Selectie van de justitiabelen

Met het oog op een effectieve werking wordt de (strafrechtelijke) gedragsinterventie toegepast op het type justitiabele waarvoor de gedragsinterventie is bedoeld. Het is belangrijk de juiste deelnemers te selecteren. Daarom moet er een duidelijke specificatie zijn van het type justitiabele waarop de gedragsinterventie zich richt en de methoden om deze doelgroep te selecteren. Een goede diagnostiek is daarbij onontbeerlijk.

Het gaat om de instroom van justitiabelen die moeten veranderen op bepaalde criminogene kenmerken en bij wie kans op recidive wordt verminderd door juist deze gedragsinterventie.

Het volgende wordt beschreven:

- de voorwaarden die gelden om deel te nemen aan deze gedragsinterventie (aard van het delictgedrag dat de gedragsinterventie beoogt te voorkomen, recidiverisico, leerstijl van de doelgroep, motivatie, persoonlijke kenmerken van de deelnemer zoals geslacht, leeftijd, etnische achtergrond)
- eventuele uitsluitingscriteria
- de selectieprocedure
- gebruikte selectiemethoden (keuze screenings- en diagnose-instrumenten, risicotaxatie; waarbij RISC (voor volwassenen) het standaardinstrument wordt geacht dat zo nodig aangevuld wordt met verdiepingsdiagnose)

⁸ Bron: <http://www.justitie.nl/onderwerpen/criminaliteit/erkenningscommissie/kwaliteitscriteria/> (gedownload 27 april 2007)

- criteria en procedures voor uitsluiting nadat het programma is gestart.

3. Gericht op dynamische criminogene factoren

De gedragsinterventie is gericht op risicofactoren die verminderd moeten en kunnen worden. Mensen die delicten plegen hebben vaak dezelfde eigenschappen of ervaringen die het risico voor recidive verhogen. Sommige van deze zogeheten criminogene factoren zijn niet veranderbaar: de statische criminogene factoren, zoals familiegeschiedenis of eerdere veroordelingen. Andere zijn wel veranderbaar. Dit zijn dynamische criminogene factoren zoals attitudes en gedrag. De gedragsinterventie dient gericht te zijn op de factoren die bij de betreffende deelnemer ook daadwerkelijk samenhangen met het criminele gedrag; op de problemen en kenmerken die het probleemgedrag veroorzaken, bevorderen of in standhouden. De gedragsinterventie beoogt dynamische criminogene factoren bij de justitiabele zelf te veranderen, maar kan tevens gericht zijn op zijn/haar omgeving (sociale netwerken, gezin).

Uit de beschrijving blijkt het volgende:

- Wat de dynamische criminogene factoren zijn waarop de gedragsinterventie zich richt. Deze risicofactoren hebben een relatie tot crimineel gedrag. Uitgelegd wordt waarom de interventie zich juist richt op deze factoren.
- Vaak is het een combinatie van factoren die de kans op recidive verhoogt. De beschrijving laat zien hoe de dynamische criminogene factoren met elkaar verbonden zijn en hoe deze factoren direct of indirect gerelateerd zijn aan de aard van het delictgedrag dat de gedragsinterventie beoogt te verminderen. Er wordt verduidelijkt dat aandacht voor deze combinatie van risicofactoren bijdraagt aan terugdringing van het recidiverisico.
- De dynamische criminogene factoren waarop de gedragsinterventie zich richt komen overeen met de bij diagnose vastgestelde risicofactoren van de deelnemende delinquenten.

4. Effectieve (behandel)methoden

Een (strafrechtelijke) gedragsinterventie heeft onder meer tot doel de dynamische criminogene factoren en andere dadertkenmerken die de kans op recidive waarschijnlijk maken, te veranderen. Hiertoe worden uiteenlopende methoden gebruikt, maar in de praktijk blijken niet alle methoden het bedoelde effect te hebben. Daarom dienen er (behandel)methoden te worden toegepast die aantoonbaar effectief of veelbelovend zijn, in die zin dat zij bijdragen aan het terugdringen van recidive.

In de beschrijving komt het volgende naar voren:

- De gedragsinterventie vindt plaats aan de hand van een duidelijke beschrijving van de gebruikte (behandel)methode. De relatie die de methode heeft tot de dynamische criminogene factoren en het type delinquent waarop de gedragsinterventie is gericht, dient theoretisch verantwoord te worden.
- Op grond van empirische gegevens wordt aangetoond dat de (behandel)methode effectief is voor de doelgroep van de strafrechtelijke gedragsinterventie.
- Daar waar de effectiviteit van de (behandel)methode nog niet empirisch is aangetoond of waar nieuwe methoden worden gebruikt, wordt aangegeven op grond van welke wetenschappelijke literatuur verwacht mag worden dat deze aanpak effectief zal zijn. Tevens dient de nieuwe aanpak toetsbaar te zijn.
- Wanneer verschillende (behandel)methoden worden gebruikt moet worden beschreven hoe deze een samenhangend geheel vormen binnen de gedragsinterventie.

5. Gericht op vaardigheden en protectieve factoren.

Binnen het kader van een (behandel)methode kunnen verschillende accenten worden gelegd. Aangetoond is dat bij zowel ambulante als intramurale behandeling/begeleiding het trainen van vaardigheden een belangrijk aspect is bij het terugdringen van recidive. Gedragsinterventies met een cognitief bestanddeel en een gedragstherapeutische basis komen als het meest effectief naar voren. Vooral de combinatie van deze twee componenten blijkt van wezenlijke invloed: het bevorderen van cognitieve vaardigheden (kennen, weten) en het inoefenen van ander gedrag.

Bij een dergelijke cognitieve gedragsaanpak zit de essentie in het leren van praktische, sociale en probleemoplossende vaardigheden. Zoals leren lezen, schrijven, werk vinden, relaties aangaan en onderhouden, en andere vaardigheden die relevant zijn in het leven van de deelnemers en een relatie hebben met hun criminele gedrag. Daarnaast wordt aandacht besteed aan het veranderen van irrationele aspecten in het denken die samenhangen met de gedragsproblemen (bijv. anderen de schuld geven), het stimuleren van de morele ontwikkeling en analyseren van de risicofactoren die samenhangen met het gepleegde delict.

De kern zit dus in het aanleren van vaardigheden. De aanpak moet concreet en praktisch zijn waarbij het belangrijk is om aan te sluiten bij de mogelijkheden die de deelnemer heeft en deze te stimuleren. Door doelen te formuleren die de deelnemer kan realiseren wordt geïnvesteerd in protectieve factoren. Ook het versterken van de binding met systemen (ouders, partner, gezin, werk, omgeving, school, etc.) maakt de kans op terugval kleiner. Het is - vooral bij jeugdigen - belangrijk om gedurende de gedragsinterventie aandacht te besteden aan deze sociale terugkoppeling. Er dient nadruk te liggen op de "community-basedness" van de aanpak, waarin ook nazorg, overdracht naar voogdij en de combinatie civiel-strafrecht een plaats krijgen.

Duidelijk wordt dat:

- De (strafrechtelijke) gedragsinterventie de protectieve factoren bevordert en de deelnemers traint in het aanleren van vaardigheden die hen kunnen helpen te leven en werken zonder opnieuw een delict te plegen, en die hen in staat stelt deel te nemen aan maatschappelijk aanvaardbare activiteiten.
- De te trainen vaardigheden een relatie hebben tot het delictgedrag. Beschreven wordt hoe de vaardigheden zijn geselecteerd (waarbij wordt aangetoond waarom deze vaardigheden belangrijk zijn voor de betreffende doelgroep of deelnemer) en hoe ze worden aangeleerd, evenals de wijze waarop de resultaten worden geëvalueerd.

6. Fasering, intensiteit en duur

Bij effectieve gedragsinterventies sluiten de intensiteit en duur van de interventie aan bij de problematiek van de deelnemer. De intensiteit en duur van de gedragsinterventie zal meestal afhankelijk zijn van de leerstijl van de deelnemer, het recidiverisico en de mate waarin de dynamische criminogene factoren meer of minder gemakkelijk te veranderen zijn.

De eerste stap is een goede diagnose bij aanvang van de strafrechtelijke gedragsinterventie. Door vervolgens gebruik te maken van verschillende, elkaar aanvullende componenten (modulen), wordt een differentiatie en fasering mogelijk die aansluit bij de behandelings-/begeleidingsbehoefte van de deelnemer.

In de beschrijving komt het volgende naar voren:

- Er wordt aangegeven hoe lang de gehele gedragsinterventie duurt en aangetoond dat deze lengte voldoende is om de gewenste veranderingsdoelen te bereiken.
- Beschreven wordt uit hoeveel sessies van hoeveel uur (elke fase van) de gedragsinterventie bestaat en wat de minimale en maximale periode is die benodigd is voor een gedragsinterventie.
- Er wordt duidelijk dat de frequentie en het aantal sessies aansluiten bij de leerstijl en -mogelijkheden van de deelnemer en dat de intensiteit (hoe vaak en binnen welke periode de sessies worden gehouden) van het programma aansluit bij het recidiverisico. Bij justitiabelen die een hoog recidiverisico hebben wordt een langere en intensievere gedragsinterventie toegepast dan bij justitiabelen met minder kans op recidive.
- Tevens wordt aangegeven wat er gebeurt wanneer er sessies of activiteiten worden gemist, als er te weinig vooruitgang wordt geboekt of wanneer er nieuwe problemen optreden.

7. Betrokkenheid en motivatie

Een gedragsinterventie zal minder succesvol zijn als de deelnemer er negatief tegenover staat. Daarom moeten gedurende de gedragsinterventie, betrokkenheid van de deelnemer bij de gedragsinterventie en motivatie voor deelname worden bevorderd en gestimuleerd.

Motivatie dient daarbij ruim te worden opgevat. In eerste instantie is het 'kunnen aangaan van een positieve werkrelatie' wellicht het hoogst haalbare voor de deelnemer. Motivatie is echter geen statische entiteit en kan zich gaandeweg ontwikkelen. De mate waarin dit gebeurt, is ondermeer afhankelijk van de manier waarop de gedragsinterventie wordt toegepast, de betrokkenheid die de begeleiders/medewerkers tonen en de mate waarin deelnemers ontvankelijk zijn voor de methode en inhoud. Een goede indicator voor betrokkenheid en motivatie is de verhouding van het aantal deelnemers dat de gedragsinterventie afrondt ten opzichte van het aantal afvallers.

In de beschrijving komt het volgende naar voren:

- De wijze waarop in de selectieprocedure de motivatie van justitiabelen voor deelname wordt vastgesteld.
- Hoe de inhoud van de gedragsinterventie en gehanteerde methoden de deelnemers motiveren tot verandering, en dat deze passen bij hun leerstijl en hun specifieke achtergrond.
- De methoden die gebruikt worden om de motivatie gedurende de gedragsinterventie in stand te houden.
- Hoe een positieve houding van de begeleiders/medewerkers tegenover de gedragsinterventie wordt bevorderd.
- Op welke wijze men denkt te komen tot een voldoende aantal deelnemers dat de gedragsinterventie afrondt.

8. Continuïteit van de gedragsinterventie en de organisaties die het aanbieden

Er moeten duidelijke verbindingen zijn tussen de betreffende gedragsinterventie en de totale aanpak van de justitiabele. De gedragsinterventie moet geïntegreerd zijn in het gehele reïntegratieplan teneinde te garanderen dat er continuïteit is tussen gedragsinterventies. Op die manier kan een soepele overgang worden bewerkstelligd tussen gedragsinterventies binnen eenzelfde organisatie, tussen gedragsinterventies door verschillende organisaties of tussen intramurale en extramurale behandeling/begeleiding. Gecoördineerde samenwerking draagt bij tot het bereiken van een maximaal effect van de behandeling/begeleiding en monitoring.

Duidelijk wordt:

- Hoe de gedragsinterventie past in het totale reïntegratieplan van de justitiabele.
- Hoe de justitiabele begeleid wordt tijdens de uitvoering van de interventie en bij de overgang van intramurale behandeling naar terugkeer in de maatschappij. De uitvoerende instanties (gevangeniswezen, justitiële jeugdinrichtingen, (jeugd)reclassering, wellicht andere organisaties) garanderen een soepele overdracht van de begeleiding en delen relevante kennis over betrokkene.
- Dat er richtlijnen zijn voor de rol van de casemanager, begeleiders en leidinggevend in dit proces; aangegeven wordt hoe casemanagers geïnformeerd worden over de doelen van de gedragsinterventie; de afspraken voor samenwerking, overdracht en communicatie tussen organisaties en medewerkers die betrokken zijn bij het resocialisatieplan van de delinquent zijn gespecificeerd.
- Op welke wijze aandacht geschonken wordt aan nazorg en aan terugvalpreventie. Met terugvalpreventie wordt bedoeld dat justitiabelen bij zichzelf de kenmerken leren te onderkennen die leiden tot het plegen van een delict, en deze leren doorbreken.

9. Interventie-integriteit: de gedragsinterventie wordt zo uitgevoerd als bedoeld is.

Er moet informatie worden verzameld over proces en inhoud om te kunnen nagaan hoe de gedragsinterventie functioneert en om de gedragsinterventie te kunnen bijsturen als die niet werkt zoals verwacht. Zonder goede monitoring weet men niet of de gedragsinterventie eigenlijk wel wordt uitgevoerd zoals het bedoeld is, met het risico dat de effectiviteit afneemt. Daarom moet er een systeem zijn dat garandeert dat de integriteit van de gedragsinterventie wordt gehandhaafd en dat afwijkingen worden gecorrigeerd. Monitoring moet een ingebouwd onderdeel zijn van de gedragsinterventie.

Er kan onderscheid worden gemaakt in monitoring van *randvoorwaardelijke of ondersteunende aspecten* (gericht op materieel, personeel en proces) en *inhoudelijke aspecten* (gericht op integriteit van de behandeling).

Met betrekking tot de *randvoorwaarden of ondersteuning* wordt beschreven:

- de beschikbare materiële en financiële middelen en voorzieningen (inclusief de omgeving; deze dient de gedragsinterventie te ondersteunen, niet tegen te werken)
- de professionaliteit van de organisatie; onder meer de managementstructuur, klinische expertise (vaardigheden en ervaringen)
- de selectie van personeel (selectieprocedure, opleiding, motivatie, vaststellen van behandelingscompetentie), training van personeel (beschrijving van training, kwaliteitsgarantie en beoordeling personeel), supervisie en ondersteuning van het personeel (beschrijving van rollen en verantwoordelijkheden van personeel, aandacht voor eventueel negatieve gevolgen van de interventie op het personeel)
- de wijze waarop continuïteit en aanwezigheid van personeel en deelnemers wordt gewaarborgd
- de wijze waarop ten behoeve van de monitoring gegevens worden geregistreerd en hoe deze informatie wordt gebruikt voor verbetering van de gedragsinterventies

- hoe de monitoring van deelnemers geschiedt door middel van daartoe geschikte instrumenten (bijv. voor- en nametingen met behulp van gevalideerde meetinstrumenten).

Met betrekking tot de *inhoudelijke aspecten of behandelingsintegriteit* wordt aangegeven:

- hoe de gedragsinterventie geïmplementeerd wordt en de wijze waarop toezicht gehouden wordt op de juiste uitvoering daarvan, zoals het op de juiste wijze gebruiken van specifieke technieken;
- hoe de behandelingsstijl van het personeel wordt gemonitord, inclusief hun houding tegenover de deelnemers;
- hoe omstandigheden of activiteiten die de behandeling zouden kunnen belemmeren worden waargenomen en aangepakt.

10. Evaluatie

Evaluatie heeft uiteindelijk tot doel na te gaan of de gedragsinterventie effectief is. Een (strafrechtelijke) gedragsinterventie is effectief, wanneer de vooraf gestelde doelen zijn bereikt. Die doelen hebben enerzijds betrekking op datgene wat men bij de delinquent denkt te bereiken (verandering van dynamische criminogene factoren, veranderde attitudes, aanleren van vaardigheden, etc.) en anderzijds op het terugdringen van de kans op terugval in crimineel gedrag. Idealiter houden deze twee factoren verband met elkaar: een effectieve behandeling kan leiden tot minder recidive. Wanneer niet vaststaat dat behandelingsdoelen gehaald worden, kan eventuele recidivevermindering niet aan de gedragsinterventie worden toegeschreven. Het vaststellen van effectiviteit in termen van behandelingsdoelen is dus van essentieel belang. Daarnaast dient enige tijd na beëindiging van de gedragsinterventie een recidivemeting te worden gedaan. De resultaten van een experimentele groep dienen vergeleken te worden met die van een controlegroep, zeker waar het gaat om nieuwe gedragsinterventies. Daarbij dienen bij voorkeur bestaande, gevalideerde instrumenten te worden gebruikt.

In de beschrijving komt het volgende naar voren:

- Voor elke (strafrechtelijke) gedragsinterventie dient een effectevaluatie gedaan te worden waarbij tenminste sprake moet zijn van een voor- en nameting en een follow-up.
- Beschreven moet worden hoe gegevens over de gedragsinterventie ten behoeve van de (periodieke en/of continue) monitoring en (zelf)evaluatie worden vastgelegd en welke instrumenten daarvoor worden gebruikt.

Bijlage 5. Standards of evidence volgens Flay et al⁹

Deze benadering hanteert een ontwikkelingsmodel, waarin ervan wordt uitgegaan dat er eerst een effectief programma wordt ontwikkeld, de effectiviteit in praktijkpilots verder wordt getoetst, om dit vervolgens als een 'evidence-based' interventie breed te implemteren. Onderstaand citaat licht toe hoe de standaard eruit ziet.

Efficacy, Effectiveness, and Dissemination

Most interventions are first evaluated by developers or others under optimal conditions, such as having ample resources and well-trained and carefully supervised intervention personnel. Yet, programs worthy of dissemination must also be effective under real-world conditions. For this reason, prevention scientists distinguish between efficacy trials and effectiveness trials (Flay, 1986). *Efficacy* refers to the beneficial effects of a program or policy under optimal conditions of delivery, whereas *effectiveness* refers to effects of a program or policy under more real-world conditions (Flay, 1986; Greenberg, 2004; Holder *et al.*, 1995, 1999; Kellam & Langevin, 2003; Last, 1988; Moscicki, 1993).

Efficacy trials require a rigorous research design, a high quality of program implementation, and researcher control over confounding factors. In an efficacy trial, for example, a researcher may test a school-based program with highly trained and supervised research staff delivering the intervention under optimal conditions. By contrast, regular classroom teachers, who have many competing demands on their time and attention every day, may be expected to deliver the intervention once it is disseminated (Hansen & Dusenbury, 2001).

Effectiveness studies focus on important factors such as the quality of implementation, which will affect program outcomes when delivered under naturalistic conditions. Furthermore, issues regarding program fidelity and adaptation as programs are "taken to scale" may contribute further variation in the expected outcomes (Elliott & Mihalic, 2004; Flay, 1986). Thus, a program that produces significant ef-

fects in an efficacy trial may or may not yield similar effects under real-world conditions.

In general, prevention research has progressed from identifying efficacious programs and policies and then testing their effectiveness in increasingly real-world conditions. Accordingly, questions of whether an intervention meets efficacy or effectiveness standards are different and may involve a different set of standards. *For a program to be found effective, it must also meet all Standards for efficacy.*

As evidence of the efficacy and effectiveness of prevention programs and policies has accumulated, the question has emerged as to when programs and policies that have evidence of both efficacy and effectiveness are appropriate for dissemination (Hansen & Dusenbury, 2001; Kellam & Langevin, 2003; Lynagh *et al.*, 2002). *A program worthy of dissemination must also meet all of the Standards for effectiveness.* Not all programs of proven effectiveness are ready for widespread dissemination. For example, a program may require special materials and special training of teachers or clinicians before it can be delivered in a way that it is effective. It is important that programs be ready for dissemination so that they can be implemented effectively, that is, in a manner that achieves the expected effects.

Thus, we outline our Standards for Evidence in three sections: efficacy, effectiveness and dissemination. Our objective in writing these standards was to articulate a set of principles for identifying prevention programs and policies that are sufficiently empirically validated to merit the labels "tested and efficacious," "tested and effective," or "tested, effective, and ready for dissemination." Consistent with SPR's mission, we were interested in prevention programs and policies that have *public health importance* (e.g. Healthy People 2010 [U.S. Department of Health and Human Services, 2000], Weissberg *et al.*, 2003). These are directed to the prevention of social, physical, and mental health problems and the promotion of health, safety, and well-being.

For establishing programs that are efficacious or effective, we emphasize research designs that can establish causal effects. That is, we want to be confident that the program or policy under question, rather than some other factor, is responsible for the observed effects. Otherwise, claims of effectiveness are likely to be biased or untrue and it is necessary to consider competing explanations that could lead to the expected outcome (Bertrand *et al.*, 2002; Holland, 1986; Manski, 1995; Rubin, 1974; Shadish *et al.*, 2002).

samhsa.gov/template.cfm?page=default, <http://www.ed.gov/about/offices/list/osedfs/index.html>?src=oc and <http://whatworks.ed.gov/>.

⁹ Flay, B.R., Biglan, A., Boruch, R.F., González castro, F., Gottfredson, D., Kellam, S., Moscicki, E.K., Schinke, S., Valentine, J.C. & Ji, P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. Prevention Science, #####

Standards		Efficacy	Effectiveness	Dissemination
<i>Specificity of the efficacy statement</i>				
1	EY 1. Statement of efficacy is of the form: Program X is efficacious for producing Y outcomes for Z population	X	X	X
	EV 1. To claim effectiveness, studies must meet all conditions of efficacy trials plus others		X	X
	DI 1. To claim readiness for dissemination, must meet all criteria for effectiveness			X
<i>Program description and measures</i>				
2	EY 2.a. Described at a level that would allow others to implement/replicate it	X	X	X
3	EV 2.a. Manuals and appropriate training and technical support readily available		X	X
4	EV 2.b. Intervention delivered under conditions expected in the real world		X	X
5	EV 2.c.i. Stated theory of causal mechanisms		X	X
6	EV 2.c.ii. Statement of "for whom?" and "under what conditions?" intervention is effective		X	X
7	BDI 2. Evidence of ability to go to scale			X
8	EY 2.b.i. The stated public health or behavioral outcome(s) of the intervention must be measured	X	X	X
9	EY 2.b.ii. There must be at least one long-term follow-up measure	X	X	X
	EY 2.c. Psychometrically sound measures			
10	EY 2.c.i. Valid measures of the targeted behavior	X	X	X
11	EY 2.c.ii. Internal consistency (alpha), test-retest reliability, and/or reliability across raters	X	X	X
12	EY 2.c.iii. At least one form of data collected by people independent of the intervention	X	X	X
	EV 2.d. Level of exposure measured in both treatment and control conditions			
13	EV 2.d.i. Integrity and level of implementation/deliver of the intervention	x	X	X
14	EV 2.d.ii. Engagement of the target audience and subgroups of interest	x	X	X
15	ODC <i>Measures of mediating variables (or immediate program effects)</i>	x	x	x
16	ODC <i>Measures of potential side-effects</i>	x	x	x
17	ODC <i>Multiple measures of constructs</i>	x	x	x
18	ODC <i>Measures of moderating variables</i>		x	x
<i>Clarity of causal inference</i>				
	EY 3. Research design allows for unambiguous causal statements			
19	EY 3.a. Design has at least one comparison condition that does not receive the tested intervention	X	X	X
20	EY 3.b. Assignment to conditions maximizes confidence in causal statements	X	X	X
	EY 3.b.i. For most kinds of interventions, random assignment (of sufficient N without sig pretest differences)			
	EY 3.b.ii. For some kinds of large-scale interventions, repeated time-series designs without randomization			
	EY 3.b.iii. Well-conducted regression-discontinuity designs (selection model is completely known)			
	EY 3.b.iv. Matched control designs with pretest equivalence and when assignment not by self-selection			
<i>Generalizability of findings</i>				
21	EY 4.a. Report specifies what/who the sample is and how it was obtained	X	X	X
22	EV 4.a. Real-world target population and the method for sampling it is explained		X	X
23	EV 4.b.i. Degree to which findings are generalizable is evaluated		X	X
24	ODC <i>Reports of subgroup analyses</i>	x	x	x
25	ODC <i>Experimental dosage studies/analyses</i>		x	x
26	ODC <i>Replication with different populations</i>		x	x
27	ODC <i>Replication with different program providers</i>		x	x
<i>Precision of outcomes</i>				
	EY 5.a. Statistical analysis allows unambiguous causal statements			
28	EY 5.a.i. Main effects analysis at the same level as the randomization and includes all cases	X	X	X
29	EY 5.a.ii. Tests of pretest differences and adjustments for them if necessary	X	X	X
30	EY 5.a.iii. When multiple outcomes are analyzed, adjustments for multiple comparisons	X	X	X
31	EY 5.a.iv. Analyses minimize possibility that effects are due to differential measurement attrition	X	X	X

32	ODC	<i>Report the extent and patterns of missing data</i>	<i>x</i>	<i>x</i>	<i>x</i>
		<i>Statistically significant effects</i>			
33	EY	5.b.i. Results reported for every measured outcome	X	X	X
34	EY	5.b.ii. A consistent pattern of statistically significant positive effects	X	X	X
35	EY	5.b.iii. No serious negative (iatrogenic) effects on important outcomes	X	X	X
36	EY	5.c. Demonstrated practical public health impact	X	X	X
37	EV	5.a. Evaluation reports some evidence of practical importance		X	X
38	BDI	3. Clear cost information readily available	<i>x</i>	<i>x</i>	X
39		<i>Report costs and cost-effectiveness analyses</i>		X	X
40	EY	5.d. Significant effects for at least one long-term follow-up	X	X	X
41	EY	5.e.i. At least 2 high-quality studies/replicates that meet all of the above criteria for efficacy	X	X ^a	X ^a
42	EY	5.e.ii. Preponderance of evidence consistent with that from the 2 highest quality studies	X	X	X
43	EV	5.b. Consistent findings from at least 2 different high-quality effectiveness trials		X	X
44	ODC	<i>The more replications the better</i>	<i>x</i>	<i>x</i>	X
45	ODC	<i>Independent replications by organizations adopting programs</i>			<i>x</i>
46	BDI	4. Monitoring and evaluation tools available to providers		X	
47	ODC	<i>Statement of factors expected to assure program sustainability</i>		<i>x</i>	

Note. Desirable standards are shown in italics and small italicized x's. Numbering in first column is simply a running count. Numbers in columns 2–4 are numbers used in the text. EY: Efficacy; EV: Effectiveness; BDI: Broad dissemination; ODC: Other desirable criteria. ^aA program may have results from effectiveness trials without separate efficacy trials.

Bijlage 6. Classificaties kwaliteit van onderzoek

Er bestaan in de literatuur verschillende classificaties om de kwaliteit van effectonderzoek, en daarmee de 'hardheid' van bewijs voor de effectiviteit van een interventie, aan te duiden. Zo hebben **West et al (2002)**¹⁰ meer dan 100 systemen voor het typeren van de kwaliteit van studies bestudeerd. Uiteindelijk hebben zij 19 systemen voor het kenmerken van de kwaliteit van de studie en 7 systemen voor het classificeren van de sterkte van het bewijs geselecteerd, die volgens hen als startpunt kunnen dienen.

Enkele die in het kader van de discussie over Erkenning van interventies in de jeugdsector zijn genoemd:

Maryland Scientific Methods scale (MSMs)¹¹:

- Niveau 1: Correlatie tussen een interventie en een afhankelijke variabele die in theorie geassocieerd is met de uitkomsten van een interventie.
- Niveau 2: De volgtijdelijkheid tussen de inzet van een interventie en de resultaten zijn duidelijk te observeren. Er is een vergelijkingsgroep zonder dat aangetoond is dat deze vergelijkbaar is met de interventiegroep.
- Niveau 3: Een vergelijking van twee of meer vergelijkbare groepen, een met en een zonder de interventie.
- Niveau 4: Als 3, maar nu met een controle over een aantal factoren om de vergelijkbaarheid van de groepen aannemelijker te maken.
- Niveau 5: A-selectie toewijzing aan vergelijkbare groepen, bestaand uit de interventie-groep en vergelijkingsgroepen.

CBO¹²

Kwaliteit studies

A1 Systematische review van tenminste twee onafhankelijk van elkaar uitgevoerde onderzoeken van A2-niveau, waarbij de resultaten van afzonderlijke onderzoeken consistent zijn.

A2 Gerandomiseerd dubbelblind vergelijkend klinisch onderzoek van goede kwaliteit van voldoende omvang

B Vergelijkend onderzoek, maar niet met alle kenmerken als genoemd onder A2 (hieronder valt ook patiënt-controle onderzoek, cohort-onderzoek)

C Niet-vergelijkend onderzoek

D Mening van deskundigen

Niveau van conclusies

1 Onderzoek van niveau A1 of tenminste 2 onafhankelijk van elkaar uitgevoerde onderzoeken van niveau A2

2 1 onderzoek van niveau A2 of tenminste 2 onafhankelijk van elkaar uitgevoerde onderzoeken van niveau B

3 1 onderzoek van niveau B of C

4 Mening van deskundigen

¹⁰ West S, King V, Carey TS, et al. (2002). Systems to Rate the Strength of Scientific Evidence. Evidence Report/Technology Assessment No. 47 (Prepared by the Research Triangle Institute-University of North Carolina Evidence-based Practice Center under Contract No. 290-97-0011). AHRQ Publication No. 02-E016. Rockville, MD: Agency for Healthcare Research and Quality. April 2002.

¹¹ Bron: Sherman, L W, Gottfredson, D, Mackenzie, D L, Eck, J, Reuter, P and Bushway, S (1997) Preventing Crime: What Works, What Doesn't, What's Promising. Report to the United States Congress, National Institute of Justice: Washington.

¹² Uit: Kwaliteitsinstituut voor de Gezondheidszorg CBO. Handleiding voor werkgroepleden, januari 2006. Bijlage A-2

American Psychological Association / APA¹³

Level 1: Well-Established Treatments

I. At least two good between group design experiments demonstrating efficacy in one or more of the following ways:

- a. superior to pill placebo, psychological placebo, or another treatment;
- b. equivalent to an already established treatment in experiments with adequate statistical power (about 30 per group).

OR

II. A large series of single case design experiments ($n > 8$) demonstrating efficacy. These experiments must have:

- a. used good experimental designs;
- b. compared the intervention to another treatment as in Ia.

AND (for both I and II):

III. Experiments must be conducted with treatment manuals.

IV. Characteristics of the clients samples must be clearly specified.

V. Effects must have been demonstrated by at least two different investigators or teams of investigators.

Level 2: Probably efficacious treatments

I. Two experiments showing the treatment is (statistically significant) superior to waiting list control group. Manuals, specification of sample, and independent investigators are not required.

OR

II. One between group design experiment with clear specification of group, use of manuals, and demonstrating efficacy by either:

- a. superior to pill placebo, psychological placebo, or another treatment;
- b. equivalent to an already established treatment in experiments with adequate statistical power (about 30 per group).

OR

III. A small series of single case design experiments ($n > 2$) with clear specification of group, use of manuals, good experimental designs, and compared the intervention to pill or psychological placebo or to another treatment.

Level 3: Possibly efficacious treatments

I. One between group design experiment with clear specification of group and treatment approach and demonstrating efficacy by either:

- a. superior to pill placebo, psychological placebo, or another treatment;
- b. equivalent to an already established treatment in experiments with adequate statistical power (about 30 per group).

OR

II. A small series of single case design experiments ($n > 2$) with clear specification of group and treatment approach, good experimental designs, at least 2 different investigators or teams, and comparison of the intervention to pill or psychological placebo or to another treatment

Level 4: Unsupported treatments

I. Treatment does not meet criteria for level 1, 2, 3, or 5.

Level 5: Possibly harmful treatments

I. At least one study demonstrating harmful effects of a treatment that otherwise would meet criteria for level 4.

¹³ Bronnen:

- Task Force on Promotion and Dissemination of Psychological Procedures, Division of Clinical Psychology, American Psychological Association. (1995). Training in dissemination of empirically-validated psychological treatments: Report and recommendations. *The Clinical Psychologist*, 48, 3-23.
- Chorpita, B.F., Yim, L.M., Donkervoet, J.C., Arensdorf, A., Amundsen, M.J., McGee, C., Serrano, A., Yates, A., Burns, J.A. & Morelli, P. (2002). Toward large-scale implementation of empirically supported treatments for children: a review and observations by the Hawaii Empirical Basis to Services Task Force. *Clinical Psychology: Science and Practice*, 9, 165-190.

Shadish et al. (2000)¹⁴

Verschillende auteurs wijzen er op dat een hiërarchie als die van de APA een oneigenlijk verschil in waardering suggereert tussen enerzijds het bewijs uit streng-gecontroleerde wetenschappelijk onderzoek en anderzijds de bewijsvoering die volgt uit praktijkmonitoring, klinische ervaring en beschrijvend klinisch onderzoek. Met name zetten de auteurs vraagtekens bij de mate waarin het bewijs uit streng wetenschappelijk gecontroleerde studies ook geldig is voor de dagelijkse praktijk van de hulpverlening.

Shadish e.a. (2000) maakten een uitvoerig classificatiesysteem waarmee effectstudies op klinische representativiteit te beoordelen zijn. Dat systeem gaat uit van een aantal aspecten waarop de interventie en de doelgroep nauwkeurig omschreven moeten zijn, wil men de representativiteit van de studie voor de praktijk kunnen bepalen. Onderstaand overzicht zet deze aspecten op een rij. Daarachter staan kenmerken van een interventie of van een doelgroep genoemd die als representatief, respectievelijk niet-representatief voor de praktijk gelden. Als een effectstudie betrekking heeft op een interventie of een doelgroep die op veel aspecten niet-representatief voor de praktijk is, zijn de uitkomsten niet zonder meer voor de praktijk geldig. Om dat te scoren, zijn door de auteurs per aspect items gemaakt.

Aspecten van beoordeling mate van representativiteit effectstudie voor de praktijk*

Aspect	Voorbeeld representatief	Voorbeeld niet-representatief
Problemen van de cliënten	De cliënten hebben een gediagnosticeerd probleem (bijv. kinderen met gedragsproblemen krijgen een training in interpersoonlijke relaties).	De problematiek van de cliënten is niet gediagnosticeerd (bijv. basisschool kinderen krijgen een training in interpersoonlijke relaties)
Heterogeniteit van de problematiek	De cliënten verschillen van elkaar qua problematiek en hebben meerdere problemen.	De problematiek van de cliënten is homogeen en enkelvoudig.
Behandelsetting	De setting biedt gewoonlijk de vorm van jeugdzorg wordt geboden die wordt onderzocht.	De setting is een wetenschappelijke onderzoekssetting die de behandeling normaal gesproken niet biedt.
Toelating	De gebruikelijke intake-route is gevolgd.	Subjecten zijn door onderzoekers geselecteerd.
Behandelaars	De behandelaars zijn praktiserende therapeuten.	De behandelaars zijn niet-clinici (bijv. Onderzoekers, studenten)
Therapie-training	De behandelaars hebben vooraf geen intensieve training in de onderzochte therapie gekregen.	De behandelaars hebben vooraf wel intensieve training in de onderzochte therapie gekregen.
Structuur van de interventie	Niet in detail gestructureerd	Gedetailleerd gestructureerd
Monitoring implementatie	Monitoring beïnvloedt de handelwijzen van de behandelaar niet.	Toepassing van de interventie sterk gestuurd door monitoring en directe feedback.
Therapie-samenstelling	De behandelaars gebruiken een mix van interventies.	De behandelaars gebruiken alleen de onderzochte techniek.
Therapie-duur	De duur van de therapie varieert onder de cliënten.	De duur van de therapie ligt voor alle subjecten vast

* Gebaseerd op Shadish e.a. (2000)

¹⁴ Shadish, W.R., Matt, G.E., Navarro, A.M. & Phillips, G. (2000). The effects of psychological therapies under clinically representative conditions: a meta-analysis. *Psychological Bulletin*, 126, 512-529. De hier gepresenteerde tekst is gebaseerd op: Yperen, T.A. van (2003). *Resultaten in de jeugdzorg: begrippen, maatstaven en methoden*. Utrecht: NIZW Jeugd.

Bijlage 7. Classificaties sterkte effect

Cohen¹⁵

De meest bekende index voor effectgrootte is Cohen's *d*. Deze maat kan zowel een negatieve als een positieve waarde hebben. Bij een positieve waarde wijst de index op een gunstig effect van de interventie, bij een negatieve waarde is er sprake van een averechts effect. In theorie kan de index elke waarde aannemen, praktisch gezien schommelt *d* meestal tussen de -2.0 en $+2.0$. Onderstaand schema biedt vuistregels om de waarde van de index te interpreteren.

<i>Een d tussen</i>	<i>Wijst op</i>
1.3 en hoger	een zeer groot effect
.80 en 1.29	een groot effect
.50 en .79	een middelgroot effect
.20 en .49	een klein effect
-.19 en .19	geen of een verwaarloosbaar effect
-.20 en -.49	een klein negatief effect
etc.	etc.

Lipsey¹⁶

Lipsey heeft 186 verschillende meta-analyses over effectieve interventies bekeken en een nauwkeurige, meer statistisch onderbouwde indeling van de *d* gemaakt. Deze indeling is te zien in onderstaande tabel. Lipsey's 'Klein', 'Middelgroot' en 'Groot' positieve effect groottes gebaseerd op 102 geselecteerde gemiddelde effectgroottes van 186 meta-analyses over Treatment effectiveness Research.

<i>Label</i>	<i>Range:</i>	<i>Midpoint</i>
Klein effect (onderste 33%)	0.00 tot 0.32	0.15
Middelgroot effect (middelste 34%)	0.33 tot 0.55	0.45
Groot effect (bovenste 33%)	0.56 tot 1.20	0.90

¹⁵ Afgeleid van: Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, p.155-159.

¹⁶ Bronnen (met dank aan Paula Speetjens):

- Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioural treatment. Confirmation from meta-analysis. *American Psychologist*, 48 (12), 1181-1209.
- Lipsey, M. W. (1990). *Design sensitivity. Statistical power for experimental research*. London: Sage Publications.

Bijlage 8. Overzicht erkenning programma's in buitenlandse databanken¹⁷

Zwikker (2006) onderzocht welke interventies die in de Databank Effectieve Jeugdinterventies als 'veelbelovend' of 'effectief' zijn aangeduid ook in buitenlandse databanken voorkomen. Deze studie bood een blik op de uiteenlopende criteria die gehanteerd worden bij de buitenlandse databanken.

			Beoordeling in buitenlandse databanken:							
Programma in Nederland:	Beoordeling in DEI¹⁸:	Buitenlands equivalent:	CASEL¹	Blueprints²	Samhsa³	NIDA⁴	PRB⁵	Promising Practices⁶	Strengthening Families⁷	Surgeon General⁸
Taakspel	effectief****	Good Behavior Game		Promising	Effective					Promising 2
Kaleidoscoop	deels effectief ^o	High / Scope Perry Preschool	Cat. 4	Promising	Model		Positief	Proven		Promising 1
VoorZorg	I.V.	Nurse Family Partnership (NFP)		Model	Model			Proven	Exemplary 2	Model 1
Triple P (Positive Parenting Program)	I.V.	Triple P					Sugg. v. positief			
Incredible Years	I.V.	Incredible Years		Model	Model			Proven	Exemplary 1	Promising 2
PMTO	I.V.	PMTO								
Minder Boos en Opstandig	deels effectief****	Coping Power (Program)			Effective	Effective				
PCIT	I.V.	PCIT								Promising 2
Levensvaardigheden	deels effectief****	Life Skills Training (?)	Cat. 4	Model	Model	Effective		Proven		Model 2
Big Brothers, Big Sisters (of R'dam) / Match	effectief ^o	Big Brothers, Big Sisters		Model	Effective			Promising / Proven		
Equip	deels effectief**	The Equip Program								
Functionele Gezinstherapie	- (geen studies)	Functional Family Therapy		Model			Positief		Exemplary 1	Model 1
Home-Start	deels effectief ^o	Home-Start					Sugg. v. positief			
PAD	effectief ^o /*	PATHS	Cat. 4	Model	Model	Effective				Promising 2
VRIENDEN	- (geen studies)	FRIENDS								
Families First	I.V.	Families First								
Youth at Risk	I.V.	Youth at Risk (YAR)								
Glenn Mills School	I.V.	Glenn Mills School								

¹⁷Bron: Zwikker, M. (2006). *Overzicht erkenning programma's in buitenlandse databanken. Interne notitie*. Utrecht: NIZW Jeugd.

¹⁸ I.V. = In Voorbereiding

Bronnen en toelichting en rating categories buitenlandse databanken

1. CASEL (www.casel.org (Publicatie: **Safe and Sound, 2003**))

Amerikaanse website met informatie over preventieve programma's op het terrein van school en sociaal-emotioneel functioneren. Biedt onder meer op onderzoek gebaseerde principes van goede programma's op dit gebied en bevat een driejaarlijkse update van de beste effectieve programma's.

Evidence of Effectiveness:

1. Program did not provide evaluations meeting specified criteria, or a preponderance of evidence does not show positive program impacts on behavioral measures
2. A single study documents positive behavioral outcomes at post-test
3. Multiple studies document positive behavioral outcomes at post-test or A single study finds positive behavioral impacts at a follow-up at least one year after the intervention ended
4. Multiple studies document positive behavioral outcomes at post-test, with at least one study indicating positive behavioral impact at follow-up at least one year after the intervention ended

2. Blueprints for Violence Prevention (www.colorado.edu/cspv/blueprints (Update 2005))

Amerikaanse informatiebron die gedegen documentatie biedt over effectieve en veelbelovende programma's, geselecteerd uit een overzichtsstudie van meer dan 600 preventieve interventies op het terrein van jeugd delinquentie, geweld en drugs.

Programs are divided into **Model** and **Promising**

As evident by its title, Blueprints for Violence Prevention's main objective is that of violence prevention in children and adolescents from birth to age 19. Programs focus on violence, delinquency, aggression (including predelinquent aggression), and substance abuse.

Criteria for Model and Promising programs include: evidence of deterrent effect with a strong research design (experimental or quasi-experimental) on one of the above outcomes. Other criteria that Model programs must meet include sustained effects for at least one year post-treatment and replication at more than one site with demonstrated effects.

3. SAMHSA / Center for Substance Abuse Prevention (CSAP), Dept. of Health & Human Services, National Registry of Effective Programs (NREP); <http://modelprograms.samhsa.gov> (Update 2005)

Uitgebreid en gedegen Amerikaans overzicht van veelbelovende en effectieve programma's voor preventie en terugdringing van risicogedrag en riskante situaties (zoals roken, gebrek aan schoolvaardigheden, veerkracht of sociale vaardigheden, gedragsproblemen, delinquentie, verslaving). Het leeuwendeel van de gepresenteerde interventies betreft kinderen, jongeren en hun ouders. Programs are divided into **Model**, **Promising**, and **Effective**

Programs are scored 1 to 5, with 1 being the lowest and 5 being the highest score, relative to 15 criteria. Model programs are well implemented and evaluated according to rigorous standards of research, scoring at least 4.0 on the 5-point scale. Promising programs have been implemented and evaluated sufficiently and are considered to be scientifically defensible, but have not yet been shown to have sufficient rigor and/or consistently positive outcomes required for Model status. Promising programs must score at least 3.33 on the 5-point scale. Effective programs meet all the criteria as the Model programs, but for a variety of reasons, these programs are not currently available to be widely disseminated to the general public.

4. National Institute of Drug Abuse (NIDA; www.nida.nih.gov/Prevention (Publicatie: **Red Book, sec. edition, 2003**))

Het National Institute on Drug Abuse in de Verenigde Staten geeft het zogeheten 'Red book' uit, een (te downloaden) gids die een overzicht geeft van op onderzoek gebaseerde preventieprogramma's voor drugsgebruik. Een deel van de programma's is gericht op preventie van gedragsproblemen en delinquentie. Ook zijn er programma's voor het aanleren van 'life skills'.

Red Book, 2003: Programs are categorized as **Effective**

No specific criteria for program inclusion mentioned.

5. PRB (Policy Research Bureau, UK; www.prb.org.uk/wwiparenting (Publicatie: **What works in parenting support, 2004**))

Het Londense Policy Research Bureau (PRB) is een organisatie die interessante overzichtsstudies en effectonderzoeken verricht op het terrein van jeugd en opvoeding. Veel rapporten zijn gratis te downloaden. Verstoopt in de website is via www.prb.org.uk/wwiparenting ('what works in parenting') een databank te vinden met details over studies naar en effectiviteit van preventieve en curatieve opvoedingsinterventies.

What works in parenting support, 2004: No rating categories mentioned. Positief beoordeelde programma's worden besproken, inclusief het verrichte evaluatieonderzoek. De bespreking geeft een idee van de mate waarin het programma als effectief beoordeeld zou worden in geval van rating. In het voorgaande overzicht van buitenlandse beoordelingen zijn deze oordelen samengevat als 'positief' of 'suggesie van positief'. Deze categorieën kunnen beschouwd worden als equivalenten van respectievelijk 'effectief' en 'veelbelovend'.

6. Promising practices (www.promisingpractices.net (Update 2005))

Deze Amerikaanse site geeft een overzicht van veelbelovende en effectieve programma's op het terrein van de gezondheid van kinderen, schoolrijpheid, schoolsucces en het gezin.

Evidence Criteria:

Type of Information	Proven Program	Promising Program	Not Listed on Site
	Program must meet all of these criteria to be listed as "Proven".	Program must meet at least all of these criteria to be listed as "Promising".	If a program meets any of these conditions, it will not be listed on the site.
Type of Outcomes Affected	Program must directly impact one of the indicators used on the site.	Program may impact an intermediary outcome for which there is evidence that it is associated with one of the PPN indicators.	Program impacts an outcome that is not related to children or their families, or for which there is little or no evidence that it is related to a PPN indicators (such as the number of applications for teaching positions).
Substantial Effect Size	At least one outcome is changed by 20%, 0.25 standard deviations, or more.	Change in outcome is more than 1%.	No outcome is changed more than 1%.
Statistical Significance	At least one outcome with a substantial effect size is statistically significant at the 5% level.	Outcome change is significant at the 10% level (marginally significant).	No outcome change is significant at less than the 10% level.
Comparison Groups	Study design uses a convincing comparison group to identify program impacts, including randomized-control trial (experimental design) or some quasi-experimental designs.	Study has a comparison group, but it may exhibit some weaknesses, e.g., the groups lack comparability on pre-existing variables or the analysis does not employ appropriate statistical controls.	Study does not use a convincing comparison group. For example, the use of before and after comparisons for the treatment group only.
Sample Size	Sample size of evaluation exceeds 30 in both the treatment and comparison groups.	Sample size of evaluation exceeds 10 in both the treatment and comparison groups.	Sample size of evaluation includes less than 10 in the treatment or comparison group.
Availability of Program Evaluation Documentation	Publicly available.	Publicly available.	Distribution is restricted, for example only to the sponsor of the evaluation.

7. Strengthening America's Families (www.strengtheningfamilies.org; 1999)

Amerikaanse site met gedegen periodieke studies en lijsten van veelbelovende en effectieve programma's, gericht op ondersteuning en behandeling van opvoeders en gezinnen ter voorkoming van delinquentie en verslaving.

Programs are divided into **Exemplary 1**, **Exemplary**, **Model**, and **Promising**. The National Program Review Committee, the University of Utah, and CSAP reviewed the programs that focused on family therapy, family skills training, in-home family support, and parenting programs. Each program was rated on theory, fidelity, sampling strategy, implementation, attrition, measures, data collection, missing data, analysis, replications, dissemination capability, cultural and age appropriateness, integrity, and program utility and placed into the following categories:

Exemplary I: Program has experimental design with randomized sample and replication by an independent investigator. Outcome data show clear evidence of program effectiveness.

Exemplary II: Program has experimental design with randomized sample. Outcome data show clear evidence of program effectiveness.

Model: Program has experimental or quasi-experimental design with few or no replications. Data may not be as strong in demonstrating program effectiveness.

Promising: Program has limited research and/or employs non-experimental designs. Data appears promising but requires confirmation using scientific techniques.

8. Surgeon General (www.surgeongeneral.gov; Publicatie: Youth Violence, 2001)

De Surgeon General is de belangrijkste functionaris in de Verenigde Staten op het gebied van gezondheidsvoorlichting en -opvoeding. Veel van de studies op de website gaan over lichamelijke gezondheid en gezondheidsproblemen, maar er is tevens een rapport te vinden over een studie naar modelprogramma's, veelbelovende programma's en interventies die niet werken in de bestrijding van geweldpleging door jeugdigen:

U.S. Department of Health and Human Services (2001). *Youth Violence. A Report of the Surgeon General*. Rockville, MD: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Injury Prevention and Control; Substance Abuse and Mental Health Services Administration, Center for Mental Health Services; and National Institutes of Health, National Institute of Mental Health.

Programs are divided into **Model** and **Promising: Level 1-Violence Prevention** and **Level 2-Risk Factor Prevention**

The primary focus of the report by the Surgeon General is violence prevention and intervention. The criteria the Surgeon General set were appropriately rigorous methods of inquiry and sufficient data to support the conclusions. **Model** programs have rigorous experimental design (experimental or quasi-experimental), significant effects on violence or serious delinquency (Level 1) or any risk factor for violence with a large effect size of .30 or greater (Level 2), replication with demonstrated effects, and sustainability of effect. **Promising** programs meet the first two criteria (although risk factors of .10 or greater are acceptable), but programs may have either replication or sustainability of effects (both not necessary).