

De kloof dichten. Een kwalitatieve basis voor erkenning van interventies.

Trudie Knijn en Anja Machielse, juli 2018.

Deel I: Achtergronden en motivatie

Introductie

Uitgaven in de sociale sector moeten steeds nadrukkelijker worden verantwoord naar overheden en financiers. Ook instellingen zelf zien het belang van effect- en evaluatieonderzoek om de kwaliteit van de hulpverlening te verbeteren (Scholte & Sprinkhuizen, 2012). Hoewel evidence-based werken in de sociale sector steeds meer erkenning vindt, is het percentage interventies in zorg en welzijn waarvan de effectiviteit 'evidence-based' is bewezen (dat wil zeggen: door middel van empirisch en theoretisch gefundeerd onderzoek) in Nederland nog steeds erg laag (Lub, De Groot & Schaafsma, 2011, Omlo, 2011).¹ Ook de feitelijke toepassing van evidence-based interventies door beroepsbeoefenaren is gering (Van der Zwet, Booijnk & Kok, 2015). Er is dus een kloof tussen de wetenschappelijke erkenning van de effectiviteit van interventies enerzijds en de praktische bruikbaarheid en toepasbaarheid van die interventies in de professionele praktijk anderzijds. Met deze notitie proberen we deze kloof te dichten. We pleiten daarbij voor een interactief model van kennisverzameling waarin de relatie tussen wetenschap en praktijk centraal staat.

De gezamenlijke kennisinstellingen in Nederland² streven naar kwaliteitsverbetering in de betrokken werkvelden door een eenduidig beoordelingssysteem voor interventies. Ze hebben erkenningscommissies aangesteld om de effectiviteit van het toenemende aantal sociale interventies te beoordelen. Deze commissies hanteren alle dezelfde werkwijze en criteria voor wetenschappelijke onderbouwing van interventies. Ze classificeren de effectiviteit van interventies op basis van de mate van robuustheid/rigiditeit van theoretische fundering en de bewijskracht van de gehanteerde methoden van onderzoek. Er worden verschillende niveaus van bewijskracht onderscheiden, oplopend van *beschrijvend* (laagste niveau) via *plausibel* en *functioneel* (middenniveau) naar *evidence-based* (hoogste niveau) (Veerman & Van Yperen, 2007). Sterke bewijskracht (*evidence-based*) vereist in dit model onderzoek met een experimentele opzet waarbij subjecten aselekt zijn toegewezen aan een onderzoeksgroep, of een andere opzet die de causale relatie tussen interventie en effect voldoende overtuigend aantoonst. Kwalitatief onderzoek is in dit model slechts mogelijk tot en met het niveau 'functioneel'.

Hoewel de erkenningscommissies ook criteria hebben opgesteld voor programma's die nog niet bewezen effectief, maar wel 'goed onderbouwd' zijn, is het streven om de 'effectiviteit' van interventies vast te stellen.³ Vooral snog gaan de erkenningscommissies uit van drie

¹ Zo is in 2007 slechts 5% van de jeugdzorginterventies evidence-based (Veerman & Van Yperen, 2007).

² RIVM-Centrum Gezond Leven (CGL), MOVISIE, het Nederlands Centrum Jeugdgezondheid (NCJ), het Nederlands Instituut voor Sport en Beweging (NISB), het Nederlands Jeugdinstituut (NJI), het Trimbos-Instituut en Vilans.

³ Dit geldt bijvoorbeeld voor 180 van de 216 interventies van het NJi.

niveaus van effectiviteit, namelijk: 1) effectiviteit volgens *eerste* aanwijzingen (op basis van observaties en casestudies), 2) effectiviteit volgens *goede* aanwijzingen (op basis van quasi-experimentele designs of time-series onderzoeken) en 3) effectiviteit volgens *sterke* aanwijzingen (aangetoond met twee Nederlandse onderzoeken met sterke bewijskracht (gerandomiseerd onderzoek met een controlegroep en zes maanden follow-up) of één Nederlands onderzoek met sterke bewijskracht in combinatie met dergelijk onderzoek uit het buitenland). In de toelichting van de criteria wordt gesproken over gecontroleerd onderzoek zoals Randomized Controlled Trials (RCT's) en quasi-experimentele studies met een follow-up, maar ook over bijvoorbeeld meerdere N=1 studies en timeseries (Erkenning van interventies, 2015).⁴

In deze notitie richten we ons op twee gesignaleerde problemen: 1) het geringe aantal interventies dat aan de gestelde criteria voldoet, en 2) de beperkte bruikbaarheid en toepasbaarheid van interventies die wetenschappelijk effectief zijn bevonden in de professionele praktijk. Onze stelling is dat het hiërarchisch model van bewezen effectiviteit waarin de toepasbaarheid van de beschreven interventie niet wordt meegenomen, negatieve effecten heeft voor het gebruik van interventies in de praktijk. In het verlengde daarvan pleiten we voor evidentie waarin theoretische en empirische rigiditeit mede is gebaseerd op toepasbaarheid in de praktijk. Om dat te bereiken is een opwaardering van kwalitatieve onderzoeksmethoden onontbeerlijk. We sluiten daarmee aan bij de notitie '*Erkenning van Interventies. Criteria voor gezamenlijke kwaliteitsbeoordeling 2015-2018*' van de gezamenlijke kennisinstituten (2015). In deze notitie stellen de kenniscentra dat kwalitatief onderzoek verder geïntegreerd dient te worden in het erkenningstraject omdat het een bijdrage kan leveren aan de onderbouwing van 'de aannemelijkheid van de werkzaamheid en inzicht kan geven in de werking en de daadwerkelijke uitvoering, en in het effect van de werkzaamheid zoals ervaren door doelgroep en uitvoerders' (Erkenning van interventies, 2015: 13). De kenniscentra hebben echter nog geen strikte criteria geformuleerd voor de kwaliteit en (bewijs)kracht van kwalitatief (evaluatie)onderzoek.

In deze notitie stellen wij voor het hiërarchische beoordelingsmodel voor de effectiviteit van interventies te herzien en wel zodanig dat *sterke aanwijzingen* voor effectiviteit gebaseerd zijn op een combinatie van theoretische robuustheid en empirisch aangetoonde bruikbaarheid in de praktijk, op basis van een evaluatie waarbij professionals en cliënten zijn betrokken. Naar onze mening kan een interventie alleen effectief zijn, als hij (ook) in de praktijk bruikbaar is. Dit betekent dat beoordeling van de effectiviteit van een interventie ook kennis over de praktische toepassingen ervan vereist. We pleiten daarom voor bewijs van effectiviteit dat theoretisch goed is onderbouwd, maar tevens is gebaseerd op werkzaamheid in de praktijk. De redenen voor die herziening worden in deze notitie toegelicht op basis van de algemeen erkende tekortkomingen van RCT's en de voorwaarden voor een praktische toepassing van interventies in het sociale domein.

⁴ Bij de beoordeling van de effectiviteit van een interventie beoordeelt de commissie eerst of de meest relevante uitkomstmaten in de studie zijn meegenomen, hoe groot de gevonden effecten zijn en tot slot het design en de kwaliteit van het onderzoek.

De consequentie van dit voorstel is tweeledig. Enerzijds is het een erkenning van de wetenschappelijke waarde van de RCT, de N=1 studies en timeseries; die kunnen bijdragen aan generaliseerbare theoretische kennis over effecten van interventies. Anderzijds is het een relativering van de bijdrage die deze kennis kan leveren aan feitelijke kwaliteitsverbetering in de werkvelden, vanwege de methodologische beperkingen en de beperkte toepasbaarheid ervan in de praktijk. Ons uitgangspunt is dat ook kwalitatieve onderzoeksmethoden (al dan niet gecombineerd met gerandomiseerd onderzoek) relevant zijn om de effectiviteit van een interventie te beoordelen omdat ze mogelijkheden bieden om de werkzame elementen van complexe interventies in kaart te brengen en daarmee de bruikbaarheid en toepasbaarheid ervan in verschillende contexten te vergroten. De keuze voor een bepaalde onderzoeksbenadering is afhankelijk van het type interventie, de ontwikkelingsfase en de complexiteit van de interventie, en de doelgroep van de interventie

Hierna formuleren we eerst de belangrijkste bezwaren tegen de eenzijdige nadruk op gerandomiseerd onderzoek om de effectiviteit van sociale interventies wetenschappelijk te bewijzen. Daarna geven we redenen om het erkenningstraject van de gezamenlijke kennisinstututen te herzien. Ook pleiten we voor een alternatieve benadering, waarin de wederzijdse beïnvloeding van wetenschap en praktijk centraal staat, en waarin kwalitatieve onderzoeksmethoden onontbeerlijk zijn.

Bezwaren tegen de eenzijdige nadruk op RCT als hoogste criterium voor wetenschappelijk bewezen effectiviteit

Tot nu toe is de Randomized Controlled Trial (RCT) een hoog geprioriteerde toets voor het wetenschappelijk bewijzen van de effectiviteit van interventies. Een zuivere RCT bestaat uit experimenteel onderzoek waarin effectmetingen plaatsvinden onder gecontroleerde condities. De interventie wordt uitgevoerd bij een *at random* samengestelde testgroep en vergeleken met een *at random* samengestelde controlegroep, die niet of op een andere manier dan de testgroep wordt behandeld. De effecten of resultaten van de interventie worden getoetst aan vooraf geformuleerde normen en het bewijs (evidence) voor effectiviteit bestaat uit scores op meetbare variabelen of gestandaardiseerde meetinstrumenten met een bekende betrouwbaarheid en validiteit (zoals gevalideerde meetschalen) (Van der Zwet et al., 2011). In deze notitie beschouwen we RCT's als onvoldoende bewijs voor het bepalen van de toepasbaarheid van een interventie in de praktijk omdat deze onvoldoende recht doen aan de complexiteit, de dynamiek en de contextuele bepaaldheid van sociale interventiepraktijken (Cartwright, 2007; Sackett, Strauss, Richardson, Rosenberg & Haynes, 2000; Tavecchio & Gerbrand, 2012).⁵

Een eerste algemeen bezwaar tegen RCT's als voldoende bewijs voor de effectiviteit van een interventie is dat de toepassing van RCT's *schijnzekerheid* biedt. In toenemende mate is

⁵ Hoewel ondertussen ook minder strikte vormen van RCT's zijn ontwikkeld en toegepast, zoals de 'pragmatische RCT (Michon & de Kroon, 2013) behandelen we die hier niet omdat deze notitie zich vooral richt op de hiërarchische positionering van de wetenschappelijk zuivere RCT ten opzichte van effectonderzoek in de praktijk.

gebleken dat RCT's te afhankelijk zijn van een contextloze setting en dat in RCT's te vaak gebruik wordt gemaakt van een homogene (blanke, hoger opgeleide middenklasse) populatie. De in RCT's gebruikte (controle)groepen weerspiegelen zelden de in de praktijk beoogde doelgroepen (die categorisch meer problematisch zijn). Bovendien zijn RCT's vaak niet herhaalbaar onder dezelfde condities en is de bewijskracht daarmee wankel (zie recente kritiek op medisch en psychologisch onderzoek, bv. Dehue, 2002; Drenth, 2016; Ioannidis, 2005; Martens, 2016). De praktische betekenis van RCT's wordt hierdoor overschat en het baseren van 'sterke aanwijzingen voor effectiviteit' op RCT's creëert een valse zekerheid.

Een tweede bezwaar tegen de veronderstelde bewijskracht via RCT's is het *praktisch onnut* ervan in de professionele praktijk (Gibbs & Gambrill, 2002; Mullen & Bacon, 2004; Teddlie & Tashakkori, 2009). Interventies die strak gereguleerd zijn en volgens voorgeschreven protocollen gehanteerd en geëvalueerd moeten worden, blijken niet geschikt te zijn voor praktisch gebruik in sociale settings (Barkham & Mellor-Clark, 2003; Durlak & Dupre, 2008; Steyaert et al., 2010; Wandersman et al., 2008). Ook professionals die bereid zijn een strak en sterk bewezen geprotocolleerde interventie te gebruiken, wijken bij de uitvoering af van de beschreven interventie. De redenen daarvoor zijn – naast tijdsgebrek, een tekort aan training, onhaalbaarheid van gestelde kwaliteitseisen - vooral gelegen in de praktische en professionele noodzaak de richtlijnen aan te passen aan de vraag van de cliënt of de doelgroep (Gordon & Cooper, 2010; Potting, Sniekers, Lamers & Reverda, 2010). Sociale professionals volgen niet nauwgezet een voorgeschreven methode, maar maken hun eigen afwegingen, afgestemd op de specifieke situatie en mogelijkheden van een cliënt of doelgroep (Bellamy, Bledsoe & Traube, 2006).

Het derde argument om RCT's als onvoldoende bewijs van effectiviteit te zien, is de veel voorkomende *multiproblematiek* en *multicausaliteit* van problemen bij de beoogde cliënten/doelgroepen, waarvoor vaak meerdere interventies tegelijk worden ingezet. Er zijn dan geen algemeen aanvaarde streefscores of *benchmarks* waaraan de resultaten van sociale interventies kunnen worden afgemeten (Potting et al., 2010). Ook functioneren cliënten/doelgroepen die hulp krijgen van sociale professionals in contexten die op allerlei manieren doorwerken en invloed hebben op de condities van de RCT's, zoals afwezigheid, motivatie, betrokkenheid, et cetera. De uitkomsten van RCT's stroken vaak niet met de doelen en prioriteiten van cliënten zelf, terwijl niet-bedoelde resultaten die voor hen wél relevant zijn onzichtbaar blijven omdat ze niet meetbaar zijn met objectieve variabelen en gestandaardiseerde meetinstrumenten (Gibbs & Gambrill, 2002; Machielse, 2017; Mullen & Bacon, 2004).⁶

Ten vierde zijn *organisatie- en beleidsdynamiek* van belang voor de adequate uitvoering van op RCT gebaseerde interventies. Organisaties die interventies inzetten, hebben hun eigen dynamiek, middelen en voorschriften die de feitelijke uitvoering van een interventie

⁶ De Groot (1994) noemt dit het dekkingsprobleem van traditionele objectieve meetinstrumenten. Te grote reductie bij het bepalen van de effecten van interventies kan leiden tot eenzijdigheid en *backlash* effecten.

beïnvloeden. Vaak ontbreken de middelen voor een volledig uitgevoerde RCT inclusief strikte rapportage, vereisen concrete situaties een verandering van aanpak, of is er geen continuïteit in de toepassing van de interventie vanwege wijzigingen in de interne verhoudingen of externe samenwerkingsverbanden. Bovendien is het stikt uitvoeren van RCT-geteste interventies op lange termijn niet haalbaar vanwege nationaal en lokaal sociaal beleid dat onderhevig is aan korte termijn-beslissingen met betrekking tot middelen, richtlijnen en doelstellingen (Anderson, Cosby, Swan, Moore & Broekhoven, 1999).

Gegeven deze bezwaren (schijnzekerheid, praktisch onnut, multiproblematiek en multicausaliteit, onzichtbaarheid van onbedoelde resultaten, en organisatie- en beleidsdynamiek) is de pretentie van wetenschappelijk bewezen effectiviteit van een interventie op basis van 'een sterke aanwijzing' die is aangetoond in een RCT zowel onterecht als een keurslijf dat de verdere ontwikkeling van effectieve interventies belemmert. De onterechte eis dat effectiviteit op een 'sterke aanwijzing' (via een RCT) gebaseerd moet zijn, kan ertoe leiden dat professionals zo'n interventie negeren omdat deze in de praktijk niet bruikbaar is en niet tot het beoogde resultaat leidt. Wat werkt in een experimentele situatie onder strikte condities en onder gelijkschakeling van contextuele omstandigheden en sociale interacties, is immers niet per se succesvol in de praktijk. Beoordeling van de effectiviteit van een interventie vereist daarom ook kennis over de praktische toepassingen ervan.

Het doel van de Databank Effectieve Interventies van MOVISIE en de andere kennisinstituten is het vergroten van de kwaliteit van de hulpverlening door een aanbod van effectieve interventies. Naar onze mening draagt dit aanbod alleen bij aan een betere situatie van de cliënten/doelgroep als het hanteerbaar en uitvoerbaar is voor professionals. Dit uitgangspunt sluit aan bij de recente wetenschapsvisie van het ministerie van OCW over principes van kennis-co-creatie, dat wil zeggen dat wetenschappers en maatschappelijke partijen samen nieuwe kennis ontwikkelen die aansluit bij de praktijk. Ze maken daarbij gebruik van de unieke kennis en vaardigheden van de eindgebruikers (OCW, 2014: 39). Deze visie van OCW is uiterst relevant voor de erkenningscommissies die de effectiviteit van sociale interventies beoordelen met het oog op innovatie van methodisch handelen. De commissies zijn zelf onderdeel van het werkveld waarin wetenschappelijke (theoretische) kennis wordt gebruikt voor praktische leer- en verbetercycli. OCW stelt dat kenniscirculatie en valorisatie niet pas aan de orde zijn als de kennis is uitontwikkeld, maar dat juist in het begin van de keten een bewustzijn nodig is van de behoefte aan kennis *bij* en van de mogelijke kennisgebruikers.

Redenen voor herziening van het erkenningstraject

In de databanken van de kennisinstituten komen slechts weinig interventies voor die voldoen aan de voorwaarden van het hoogste niveau van effectiviteit, namelijk bewezen effectiviteit door *sterke* aanwijzingen uit een gerandomiseerd onderzoek met een controlegroep. Diverse studies over de praktische relevantie van wetenschappelijk bewezen interventies, zowel in Nederland als internationaal, laten bovendien zien dat

wetenschappelijk geteste interventies weinig worden toegepast in de praktijk (Barkham & Mellor-Clark, 2003; Durlak & Dupre, 2008; Wandersman, Duffy, Flashpohler, Noonan, Lubell, Stillman & Saul, 2008). MOVISIE heeft onderzocht waarom professionals, welzijnsorganisaties en gemeenten zo terughoudend zijn om wetenschappelijk bewezen interventies, zoals die in de databank van MOVISIE en andere kennisinstituten zijn opgenomen, toe te passen of aan te bevelen als voorwaarde voor subsidiëring (Van der Zwet et al., 2015). Dit onderzoek maakt duidelijk dat beleidmakers en bestuurders geen evidence-based interventies voorschrijven omdat ze niet in de schoenen van het uitvoerend werk willen gaan staan; vanwege onbekendheid met de ins en outs van de kennis die via de databanken wordt verschaft, of omdat er geen behoefte is aan wijziging van organisatorische systemen die hun waarde hebben bewezen. Professionals zijn terughoudend vanwege een groot vertrouwen in hun eigen expertise, een gebrek aan discretionaire ruimte en een hoge werkdruk. Ook zien professionals in de praktijk weinig toegevoegde waarde van interventies die weliswaar wetenschappelijk zijn getest, maar in de praktijk hun werkzaamheid niet hebben bewezen (Van der Zwet et al., 2015).

De centrale vraag van MOVISIE en de andere kennisinstituten is *of* en *hoe* professionals in het sociale domein meer en beter gebruik kunnen maken van wetenschappelijk effectief bewezen interventies die ook in de praktijk hun meerwaarde hebben bewezen, op welke wijze die meerwaarde door wetenschappelijk onderzoek kan worden aangetoond en, meer in het bijzonder, op welke wijze kwalitatieve onderzoeksmethoden daarbij behulpzaam kunnen zijn. Belangrijke stakeholders bij het beantwoorden van deze vraag zijn de beroepsbeoefenaren; de sociaal werkers die dag in dag uit interventies plegen en de voornaamste schakel zijn tussen beleidsvoornemens, beleidsprogramma's en hulpverlening aan de cliënten en doelgroepen die gewild of ongewild onderwerp zijn van beleid en interventie in hun eigen of het algemeen belang. Cruciaal is dat deze beroepsbeoefenaren zich kunnen committeren aan de voorgestelde interventies omdat ze overtuigd zijn van de werkzaamheid ervan, en niet omdat de interventie van bovenaf wordt opgelegd vanwege een wetenschappelijk bewijs. Alleen als professionals betrokken zijn bij de ontwikkeling, het uitvoeren en het evalueren van een interventie krijgen zij inzicht en vertrouwen in de werking ervan, en in het effect zoals zijzelf en de doelgroep dat ervaren en waarnemen (Davies & Gray, 2016; Gray, Sharland, Heinsch & Schubert, 2015). Een mooi voorbeeld daarvan is de interventie *Thuis op Straat*, een effectief bewezen interventie, die in samenspraak met professionals en cliënten wordt verbeterd op basis van praktijkevaluaties.

Een manier om de relatie tussen wetenschappelijk getoetste kennis en de toepassing ervan in de professionele praktijk te bestendigen, is het onderscheiden van *fidelity* (trouw aan RCT-based intervention) en *adaptation* (aanpassing aan de praktijk) (Durlak & Dupre, 2008). Volgens Durlak en Dupre verliezen de meeste RCT's in de praktijk hun robuustheid ('rigour') vanwege noodzakelijke aanpassingen. Wandersman et al (2008) onderscheiden verschillende fasen in het toepassingsproces van een wetenschappelijk bewezen interventie, in dit geval een preventieve interventie: Prevention Translation Systems (vertaling en verspreiding van kennis en informatie), Prevention Support Systems (training van professionals, ondersteuning bij implementatie van de interventie) en Prevention Delivery Systems (de feitelijke uitvoering van de interventie). In elke fase is consultatie van

professionals nodig, en hun ervaringen hebben weer gevolgen voor de vormgeving van de interventie (bijstelling, accenten leggen, instructies herzien, doelgroep beter afbakenen, et cetera) (Soydan & Palinkas, 2016). Dit alles hoeft geen probleem te zijn voor een effectmeting, zolang de kernelementen van de interventie, dat wil zeggen de *werkzame bestanddelen*, gehandhaafd blijven.

Om de effectiviteit van een interventie te beoordelen, moet allereerst worden vastgesteld wat de werkzame bestanddelen of kernelementen van de interventie zijn. Een praktijkgerichte evaluatie moet vervolgens uitwijzen hoe de interventie in de professionele praktijk wordt geïmplementeerd: 1) wordt de interventie volledig uitgevoerd, 2) worden de kernelementen goed uitgevoerd, 3) bij welke deelnemers, hoe betrokken zijn ze, wat is hun motivatie, 4) hoe onderscheidt de interventie zich van andere interventies, 5) aan welke andere activiteiten doen de deelnemers nog meer mee, et cetera (Durlak & Dupre, 2008)? Voorwaarde voor een praktijkgerichte evaluatie is dat professionals eraan meewerken door het verrichten van vergelijkende casestudies, het signaleren van problemen bij de implementatie en de variatie in effect bij verschillende cliënten/doelgroepen, en het onderzoeken van verschillende effecten en benaderingen voor cliënten. De resultaten van dergelijk onderzoek hebben gevolgen voor het ontwerp van de interventie, door een proces van feedback en wederzijdse beïnvloeding ('case based learning'). De (empirische) cirkel kan ook beginnen bij practice-based interventies, van waaruit geabstraheerd wordt naar theoretische assumpties, die vervolgens empirisch worden getoetst om zo de interventie te verbeteren. Het gaat erom te achterhalen onder welke omstandigheden en voor wie de interventie werkt (De Vlaeminck, 2005).

Een alternatief: de transdisciplinaire benadering

Onderzoek in de interventiepraktijk waarbij beroepsbeoefenaren zijn betrokken, kan uitwijzen of een interventie daadwerkelijk effectief is (Barkham & Mellor-Clark, 2003; Durlak & Dupre, 2008). We pleiten daarom voor een interactief en complementair systeem waarin interventies pas als effectief worden erkend als ze ook in de praktijk van het professionele werk hun waarde hebben bewezen. Het basisprincipe van dit systeem is wederzijdse beïnvloeding tussen wetenschap en (beroeps)praktijk, oftewel 'transdisciplinaire' kennisverzameling, waarbij verschillende soorten en niveaus van kennis elkaar aanvullen (Moulaert, MacCallum, Mehmood & Hamdouch, 2013). Zo'n transdisciplinaire benadering veronderstelt overeenstemming over de probleemdefinitie en de te volgen route (interventie/methodiek) naar het gewenste resultaat tussen onderzoekers, professionals en gebruikers. Dit vereist een niet-hiërarchische samenwerking tussen deze partijen, ondersteund door kennisinstituten en organisaties (werkplaatsen) voor toegepast onderzoek. Theoretische kennis wordt gevalideerd binnen een inter- en transdisciplinaire context, waarin onderzoekers samenwerken met professionals en cliënten, om de professionele praktijk te verbeteren (Gibbons, Limoges, Nowotny, Schwartzmann, Scott & Trow, 1995; Moulaert et al., 2013). Evaluaties door middel van kwalitatief onderzoek zijn daarbij van fundamenteel belang. Ze leveren kennis die de aannemelijkheid van de werkzaamheid van een interventie kan onderbouwen en vormen een belangrijke aanvulling

en/of alternatief voor op RCT's gebaseerd onderzoek. Bovendien leveren ze informatie op die de bruikbaarheid van de interventie kan vergroten, namelijk over de effectiviteit van het uitvoeringsproces (*implementatie-effectiviteit*) en de ervaren effecten van de interventie door personen uit de doelgroep zelf (*experientiële effectiviteit*) (Davies, 2005). De implementatie-effectiviteit wordt in beeld gebracht door professionals te vragen hoe de begeleidingstrajecten met cliënten verlopen. De ervaren ('experientiële') effectiviteit wordt in beeld gebracht door aan cliënten te vragen wat hun problemen en vragen waren bij aanvang van de interventie en in hoeverre ze ten aanzien daarvan baat van de interventie hebben ondervonden (Davies, 2005). Alleen door een combinatie van wetenschappelijke kennis waarin theorievorming centraal staat (het *wat*) met kennis over praktische toepassingen van verworven kennis (het *hoe*), kan innovatie van het methodisch handelen binnen een professionele context plaatsvinden.

In deze transdisciplinaire benadering wordt praktijkkennis dus nadrukkelijk verondersteld bij het vaststellen van de effectiviteit van een interventie. Uitgangspunt is dat wetenschappelijke kennis alleen onvoldoende basis biedt voor een beslissing over de beste en dus meest effectieve interventie; ook de professionele expertise en de voorkeuren en omstandigheden van de cliënt worden meegewogen bij het vaststellen van de effectiviteit (McNeece & Thyer, 2004; Sackett et al., 2000). Evidence-based werken wordt hier opgevat als een proces waarbij de professional bepaalt welke interventie wordt toegepast, op basis van het best beschikbare wetenschappelijke bewijs, de eigen expertise, en de voorkeuren, ervaringen en evaluaties van de cliënt. Vaak is er sprake van indirect bewijs of 'circumstantial evidence': het cruciale bewijs is niet hard, maar vormt in combinatie met andere aanwijzingen het uiteindelijke oordeel over het effect van de interventie (Van der Laan, 2003).

De waarde van kwalitatief onderzoek

Deze benadering heeft gevolgen voor de hiërarchie in de classificatie van interventies. We gaan ervan uit dat een interventie pas bewezen is als deze in de praktijk is getoetst bij verschillende doelgroepen en in verschillende contexten op basis van een transdisciplinaire benadering, en zien de betrokkenheid en inbreng van beroepsbeoefenaren die de interventies implementeren als voorwaarde voor de bewijskracht van interventies. Dit vraagt om een interactief model waarin de wederzijdse beïnvloeding en samenwerking van wetenschap en praktijk centraal staat. De transdisciplinaire benadering biedt hier mogelijkheden voor. Zij kan de complexiteit van de onderzochte werkelijkheid in beeld brengen door niet alleen waarde toe te kennen aan theoretische kennis over (wetenschappelijk bewezen) effectieve interventies, maar ook aan de vaak impliciete (taciëte) kennis en expertise van professionals, en de preferenties, behoeften en waarden van cliënten. Deze soorten kennis zijn noodzakelijk om de professionaliteit in de sociale sector te vergroten. Door deze benadering kunnen werkzame bestanddelen van bestaande interventies beter worden benoemd en geëvalueerd. Dit kan de kloof tussen wetenschap en praktijk dichten en tot een kwaliteitsverbetering in de sociale sector leiden.

Dit alles heeft consequenties voor de erkenningstrajecten van de kennisinstituten en de waardering van de verschillende onderzoeksmethoden om de effectiviteit van interventies te evalueren. Naast kwantitatief en experimenteel onderzoek kan kwalitatief onderzoek naar sociale interventies hieraan een belangrijke bijdrage leveren. Dat type onderzoek biedt inzicht in de omstandigheden van personen in de doelgroep, de knelpunten en de succesfactoren van interventies, en de mogelijkheden en grenzen van de hulpverlening. Op basis daarvan kunnen professionals betere diagnoses stellen en zijn ze beter in staat om realistische en haalbare doelen te formuleren. Deze kennis vormt tevens de grondslag voor het nemen van weloverwogen beslissingen over de interventies die ze inzetten bij cliënten in vergelijkbare situaties en leidt daarmee tot praktijkinnovatie en kwaliteitsbevordering. Dit veronderstelt echter ook dat heldere criteria worden gesteld voor kwalitatieve onderzoekdesigns, om de betrouwbaarheid en validiteit ervan te optimaliseren. Dit geldt des te meer wanneer de data zijn bedoeld om de effectiviteit van interventies te bewijzen aan beleidsmakers en sociale professionals en daarmee de bruikbaarheid ervan voor beleid en praktijk te vergroten (Tavecchio & Gerrebrand, 2012).

Deel II: Onderzoek naar effectiviteit: criteria voor empirisch onderzoek, kwantitatief en kwalitatief⁷

De erkenningscommissies van de verschillende kennisinstituten hebben aangegeven behoefte te hebben aan evaluatiecriteria voor bewezen effectiviteit van interventies op basis van kwalitatief onderzoek. Feitelijk gaat het hier om twee te onderscheiden aspecten: 1) de wetenschappelijke onderbouwing van de interventie en 2) de evaluatie van de interventie. In de criteria voor de bewezen effectiviteit van de interventie moeten deze aspecten zowel afzonderlijk als in samenhang worden verhelderd. Als de bewijskracht van kwalitatief evaluatieonderzoek het niveau van een *sterke aanwijzing* wil behalen, dan moet aan de volgende criteria (T4) worden voldaan:

1. **T**heoretische onderbouwing van de interventie en de daarin veronderstelde werkzame mechanismen, op basis van een passende literatuurverkenning, en de *evaluatie*.
2. **T**ransdisciplinaire samenwerking met professionals en cliënten bij het ontwerp van de *interventie*, of de aanpassing van een elders verrichtte interventie, de formulering van de *evaluatievraag*, de opzet van het *evaluatie*onderzoek, en de selectie van de steekproef.
3. **T**ransparant, betrouwbaar en valide onderzoeksdesign van de *evaluatie* (steekproef, dataverzameling en data-analyse, met gebruikmaking van daarvoor geëigende analysemethoden zoals Atlas.ti 6.0 (www.atlasti.com), MaxQDA12 (www.maxqda.com) of QSR Nvivo (www.qsrinternational.com))

⁷ De hier geformuleerde criteria voor kwalitatief onderzoek geven mogelijk aanleiding ook de criteria voor evaluatieonderzoek op basis van RCT's of grootschalig surveyonderzoek te herzien; daarover worden in deze notitie geen uitspraken gedaan.

4. **T**erugkoppeling gedurende de *evaluatie* naar uitvoerenden en cliënten en verantwoording van eventuele aanpassingen in relatie tot de veronderstelde werkzame mechanismen.

Ad 1. Theoretische onderbouwing van evaluatieonderzoek

Theoretische onderbouwing van evaluatieonderzoek blijkt een lastige opgave te zijn. Zo laat Lub (2017) zien dat aan een serie van interventies gericht op maatschappelijke activering alleen in uitzonderingsgevallen een coherente theorie ten grondslag ligt. In veel gevallen is de interventie onderbouwd met een waaier van niet altijd samenhangende theoretische noties (ook wel *theorettes* genoemd) die bovendien weinig relatie vertonen met de gehanteerde interventiemethode. Daarnaast signaleert Lub onzorgvuldig gebruik van theorie in de vorm van ‘window dressing’, tautologisch redeneren of verwarring van theorie en stappenplan.

Wat geldt voor onderzoek naar maatschappelijke activering zal vermoedelijk ook gelden voor veel andere interventies in het sociale domein, denk bijvoorbeeld aan onderzoek naar participatie, zelfredzaamheid, eenzaamheid, sociale cohesie in de wijk of begeleid wonen. Voor evaluatieonderzoek van sociale interventies impliceert dit dat allereerst kritisch geëvalueerd moet worden wat de theoretische assumpties van de interventie zijn. Pas dan kan een theorie over het effect van de interventie ontwikkeld worden, dat wil zeggen *een logisch samenhangend stelsel van assumpties op verschillende abstractieniveaus over de redenen waarom de interventie een bepaalde uitkomst zal opleveren bij deze specifieke persoon/groep/wijk onder welke condities (zie ook Swanborn, 2004: 138; Lub, 2017: 10)*. Die theorie geeft richting aan het onderzoek naar de vermoedelijk werkzame bestanddelen van de interventie. In alle gevallen is het bij evaluatieonderzoek van belang dat vooraf wordt bepaald wat de aard van de evaluatie zal zijn. Daarbij moet onderscheid gemaakt worden tussen een *planevaluatie*, gericht op het plan van aanpak dat voorafgaat aan de uitvoering van de interventie, *productevaluatie*, gericht op de uitkomsten van de interventie en *procesevaluatie*, gericht op de implementatie van de interventie (Tavecchio & Gerrebrands, 2012). Een ander bruikbaar kader is de evaluatietypologie van Owen waarin vijf hoofdvormen van evaluatieonderzoek worden onderscheiden, waarbij de afstemming van de vraagstelling op het stadium waarin de interventie of methodiek zich bevindt centraal staat (Owen, 2007).

Ten slotte een voetnoot bij de veronderstelling dat op basis van kwalitatieve casestudies geen generaliseerbare conclusies kunnen worden getrokken. Een relevante term in dit verband is *inference transferability*, de overdraagbaarheid van conclusies naar vergelijkbare casussen (Teddlie & Tashakkori, 2003). Zeker als gebruik gemaakt wordt van vergelijkende casestudies, bijvoorbeeld de ‘fuzzy set’ methode die ontwikkeld is door Ragin (2000), of de Context, Mechanism, Outcome (CMO) methode (Pawson & Tilley, 1997) bieden casestudies uitermate betrouwbare resultaten. Zulke methoden maken het mogelijk individuele (gedrags-)kenmerken in combinatie met contextuele,

institutionele, beleidsmatige en organisatorische factoren te analyseren, een voorwaarde waar veel andere onderzoeksmethoden (RCT, grootschalige surveys) niet aan kunnen voldoen. In aanvulling daarop stelt Flyvbjerg dat casestudies een ideale opzet bieden voor 'falsificeerbaarheid' van algemene theoretische veronderstellingen (het Popperiaans axioma).

One can often generalize on the basis of a single case, and the case study may be central to scientific development via generalization as supplement or alternative to other methods. But formal generalization is overvalued as a source of scientific development, whereas "the force of example" is underestimated. (Flyvbjerg, 2004: 228)

Criteria voor theoretische onderbouwing van de interventie en de evaluatie

- Er is een heldere en op theoretische en empirische literatuur gebaseerde samenvatting van de theoretische assumpties die ten grondslag liggen aan de interventie, en haar werkzame mechanismen.
- Er is een heldere beschrijving van de inhoud en de aard van het evaluatieonderzoek. Beschreven wordt welke van de theorie van de interventie afgeleide werkzame mechanismen worden geëvalueerd, en of het een plan, een product of een procesevaluatie betreft.

Ad 2. Transdisciplinaire samenwerking met professionals en cliënten

Voor kwalitatief onderzoek dat de bruikbaarheid van de interventie in de praktijk moet evalueren, geldt dat transdisciplinariteit onvermijdelijk is. Niet alleen omdat professionals de interventie in praktijk brengen maar vooral omdat zij en de cliënten context-afhankelijke kennis bezitten die de meeste sociale wetenschappers ontberen. Nogmaals Flyvbjerg (2006: 223):

First, it is important for the development of a nuanced view of reality, including the view that human behavior cannot be meaningfully understood as simply the rule governed acts found at the lowest levels of the learning process and in much theory. [.....] Concrete experiences can be achieved via continued proximity to the studied reality and via feedback from those under study. [...]

Om de praktijkwaarde van kwalitatief onderzoek te vergroten, blijft het noodzakelijk – naast wetenschappelijke vormen van kennis – ook context-afhankelijke kennis van de professionals en de cliënten in verschillende fasen van een interventie te benutten (Major et al., 2011). Dat geldt voor nagenoeg alle fasen van het onderzoek; zowel bij het ontwerp van een interventie of de aanpassing van een elders verrichtte interventie, als bij de formulering van de evaluatievraag, de opzet van het evaluatieonderzoek, en de selectie van de steekproef.

Criteria voor transdisciplinaire samenwerking

- Beschreven wordt wat *inhoudelijk* de bijdrage is geweest van de wetenschap, de professionele praktijk en de doelgroep aan de formulering van de onderzoeksvraag, de onderzoeksopzet en de selectie van de doelgroep. Er wordt uitgelegd met welke context-gebonden factoren rekening is gehouden bij de vertaling van theoretische assumpties naar de opzet van het evaluatieonderzoek, en wat de resultaten van die vertaling zijn.
- Beschreven wordt welke stappen zijn gezet om transdisciplinaire kennis te ontwikkelen; wie, wat en op welk moment betrokken is geweest in het proces.

Ad 3. Transparant, betrouwbaar en valide onderzoeksdesign voor de kwalitatieve evaluatie

Voor kwalitatief evaluatieonderzoek zijn veel verschillende onderzoeksopzetten beschikbaar. Welke onderzoeksopzet wordt gekozen, is afhankelijk van de vraagstelling, de doelgroep, de context en de reikwijdte en de fase van de implementatie. De MOVISIE-publicatie 'Wat Werkt' (2016), Swanborn's boek 'Evalueren' (2004) en Owen's boek 'Program Evaluation' (2007) en Lub's boek 'Kwalitatief evalueren in het sociale domein' (2014) geven overzichten van die variëteit, die niet eenvoudig zijn samen te vatten. Kwalitatieve onderzoeksmethoden kunnen de vorm aannemen van focusgroepen, de Delphi-methode hanteren, bestaan uit enkele of vergelijkende casestudies, of simpele tellingen combineren met diepte-interviews. Voor de beoordeling van de kwaliteit, de betrouwbaarheid en de validiteit van al die soorten kwalitatief onderzoek gelden in elk geval de volgende criteria:

Criteria voor Transparantie van kwalitatief evaluatieonderzoek

Verantwoording van de onderzoeksgroep. Daaruit blijkt het volgende:

- De deelnemers representeren de doelgroep van de interventie (demografische gegevens, specifieke kenmerken doelgroep).
- Het aantal en de variatie van de deelnemers is voldoende om gefundeerde conclusies te kunnen trekken over de doelgroep.
- In welke andere interventies de deelnemers participeren, en of en hoe dat de resultaten van de interventie beïnvloedt.
- Of de deelnemers at random of op basis van specifieke kenmerken zijn geselecteerd, en in het laatste geval: welke kenmerken dat zijn.
- Als er sprake is van een controlegroep geldt bovenstaande ook voor de controlegroep.

Verantwoording van de onderzoeksopzet

- De onderzoeksopzet bevat een beschrijving van de contextuele en organisatorische mogelijkheden en restricties van zowel de interventie als de evaluatie.
- Het gehanteerde onderzoeksdesign (bijvoorbeeld een vergelijkende casestudy), de dataverzamelingmethoden (variërend van documenten, interviews, tot focusgroepen of cijfermatige verslagen) en de data-analyse zijn geschikt om de theoretische assumpties (over werkzame mechanismen) te onderzoeken die ten grondslag liggen aan de interventie.
- Er wordt beschreven of het gaat om planevaluatie, productevaluatie of procesevaluatie, en de aard van de evaluatie is in overeenstemming met de theoretische assumpties van de interventie.
- De te verwachten uitkomsten zijn beschreven in overeenstemming met de theoretische assumpties van de interventie.
- Bij een product- en procesevaluatie heeft een nulmeting plaatsgevonden waarvan de resultaten zijn beschreven, deze worden vergeleken met de resultaten van de interventie. Bij een planevaluatie wordt beschreven wat de aanleiding is om een interventie te ontwikkelen of in te zetten (de beginsituatie), op een manier die vergelijking met de resultaten van de interventie mogelijk maakt.
- De operationalisering van de theoretische assumpties over veronderstelde werkzame mechanismen is adequaat beschreven en de dataverzamelingmethoden (zoals documenten, interviews, cijfermatige verslagen van effecten op wijk-/buurniveau of metingen van veranderingen in individueel gedrag) zijn geschikt om de theoretische assumpties die aan de interventie ten grondslag liggen te onderzoeken.

Data-analyse

- De verzamelde data zijn betrouwbaar, valide, controleerbaar en herhaalbaar. Hierbij moet de kanttekening worden gemaakt dat de geschiedenis zich nooit herhaalt en dat het onmogelijk is in een zich ontwikkelende samenleving eenzelfde situatie opnieuw te creëren.
- Er is gebruik gemaakt van analyse-instrumenten zoals Atlas ti 6.0 (www.atlasti.com), Max QDA (www.maxqda.com) of QSR Nvivo (www.qsrinternational.com) zodat gecontroleerd kan worden hoe adequaat kwalitatieve data zijn geanalyseerd in relatie tot de theoretische assumpties.
- Uit de resultaatbeschrijving blijkt welke assumpties van de interventie wel en niet bevestigd worden door het evaluatieonderzoek.
- De conclusies zijn overdraagbaar naar andere – vergelijkbare - situaties en casussen die geen deel uitmaakten van het evaluatieonderzoek.

Dataopslag (wettelijk verplicht vanaf 28 mei, 2018)

- De dataopslag voldoet aan de EU-criteria daarvoor, dat wil zeggen dat er een Privacy Impact Assessment (PIA) is gemaakt en dat de persoonsgegevens en de resultaten van de studie gescheiden en versleuteld zijn opgeslagen.

Ad 4. Terugkoppeling gedurende de evaluatie

- De rapportage maakt duidelijk op welke wijze tussentijdse bevindingen en eventuele aanpassingen in de theoretische assumpties gedurende de evaluatie zijn teruggekoppeld naar uitvoerende professionals en cliënten. Ook wordt beschreven *of* en *hoe* uitvoerenden en cliënten instemmen met de conclusies van de evaluatie en met eventuele aanpassingen ten aanzien van de veronderstelde theoretische assumpties omtrent werkzame mechanismen.

Deel III: Herziening van de indeling in ‘Erkenning van Interventies. Criteria voor gezamenlijke kwaliteitsbeoordeling 2015-2018’.⁸

Algemene criteria voor effectiviteit

Voor alle niveaus van effectiviteit geldt:

- De gevonden uitkomsten zijn (de meest) relevant(e) gegeven het doel en de doelgroep van de interventie.
- De interventies zijn gebaseerd op de meest recente theorievorming over werkzame mechanismen en zijn ingebed in een overzicht van recente empirische studies met betrekking tot het onderzoeksdomein.
- De veranderingen hebben betrekking op het doel en de doelgroep van de interventie.
 - o De studies laten zien dat de beoogde doelgroep daadwerkelijk is bereikt.
 - o De studies laten zien dat de interventie bruikbaar is voor beroepsbeoefenaren in de praktijk.
 - o De gehanteerde instrumenten bieden een betrouwbare en valide operationalisering van centrale concepten (de kernelementen van de interventie) om de realisatie van de doelen van de interventie te meten.
 - o Er zijn adequate onderzoekstechnieken toegepast.
- Bij kwantitatief onderzoek is de grootte van het effect weergegeven in een geaccepteerde uitkomstmaat, redelijk overtuigend en passend bij de aard, het doel en de doelgroep van de interventie.
- Bij kwalitatief onderzoek is het verschil tussen de beginsituatie en de eindsituatie van een interventie op overtuigende wijze weergegeven, rekening houdend met de aard, het doel, de doelgroep en de context van de interventie.

⁸ Erkenning van interventies. Criteria voor gezamenlijke kwaliteitsbeoordeling. Utrecht, 2015.

- Er is duidelijk verwoord dat zowel het uitvoeringsproces van de interventie als de uitkomstmaat is geëvalueerd door de doelgroep en andere stakeholders, waaronder de professionals.
- Eventuele negatieve effecten zijn weergegeven.
- Het onderzoek is zodanig gedocumenteerd, dat replicatie van de studie mogelijk is, alhoewel daarbij rekening moet worden gehouden met specifieke kenmerken van de nieuwe doelgroep, een andere of gewijzigde context en mogelijk verschillende condities.
- De interventie is uitgevoerd zoals bedoeld. Aangetoond is dat de werkzame elementen van de interventie daadwerkelijk zijn toegepast.
- Er zijn naar het oordeel van de commissie voldoende studies waaruit is gebleken dat er bij de uitvoering van de interventie veranderingen zijn opgetreden conform het doel van de interventie.

1. Eerste aanwijzingen voor effectiviteit

- Zie de algemene criteria voor effectiviteit.
- De opzet van het kwalitatief of kwantitatief onderzoek voorziet in een minstens lichte bewijskracht. Er zijn resultaten beschikbaar voorafgaand, bij de start of tijdens de interventie en na afloop van de interventie (zgn. voor- en nameting), met of zonder controle-conditie. Bij onderzoek naar de ervaren effectiviteit volstaan de resultaten van een nameting.
- Er zijn minstens twee Nederlandse studies naar de voorliggende interventie met een lichte bewijskracht of één Nederlandse studie naar de voorliggende interventie in combinatie met minstens één (inter)nationale studie naar deze of een vergelijkbare interventie met minstens lichte bewijskracht. In het laatste geval is aannemelijk gemaakt dat het doel, de doelgroep, de aanpak en het theoretische model van de vergelijkbare interventie voldoende van toepassing zijn op de Nederlandse situatie en/of de voorliggende interventie. De erkenningscommissie oordeelt uiteindelijk over de vergelijkbaarheid.
- Als gebruik wordt gemaakt van een (quasi-)experimenteel empirisch onderzoek dan moet daarbij precies worden toegelicht voor welk afgebakend probleem de opzet een oplossing kan bieden, welke specifieke doelgroep daarmee bereikt kan worden, wat de bestaande situatie is waarin de interventie kan worden toegepast, wat de gewenste situatie is en hoe beroepsbeoefenaren de interventie kunnen gebruiken.

2. Goede aanwijzingen voor effectiviteit

- Zie de algemene criteria voor effectiviteit.
- De opzet van het empirisch onderzoek voorziet in een minstens redelijke (causale) bewijskracht. Het onderzoek heeft een (quasi-)experimentele opzet of bestaat uit een

(vergelijkende) casestudie. De studies zijn tenminste eenmalig in de praktijk uitgevoerd of hebben nog geen follow-up.

- Het aantal studies kan sterk uiteenlopen, afhankelijk van de kwaliteit en de aard van de studie. Vuistregels voor het minimum zijn:
 - o Er zijn minstens twee Nederlandse studies naar de voorliggende interventie met een redelijke tot vrij sterke bewijskracht of één Nederlandse studie naar de voorliggende interventie in combinatie met minstens één (inter)nationale studie naar deze of een vergelijkbare interventie met minstens redelijke bewijskracht. In het laatste geval is aannemelijk gemaakt dat het doel, de doelgroep, de aanpak en het theoretische model van de vergelijkbare interventie voldoende van toepassing zijn op de Nederlandse situatie en/of de voorliggende interventie. De erkenningscommissie oordeelt uiteindelijk over de vergelijkbaarheid.
 - o Bij Nederlands onderzoek naar de voorliggende interventie met een sterke tot zeer sterke bewijskracht volstaat één studie voor de erkenning op dit niveau van effectiviteit.
 - o Bij herhaalde of vergelijkende casestudies zijn er minstens 3 cases uitgevoerd door verschillende behandelaars in verschillende condities of minstens 5 cases uitgevoerd in één setting.
- Het onderzoek heeft voldoende aannemelijk gemaakt dat 1) de interventie volledig is uitgevoerd, 2) de kernelementen ook goed zijn uitgevoerd, 3) de deelnemers de doelgroep goed representeren, 4) hoe het programma zich onderscheidt van andere programma's, 5) en aan welke andere activiteiten de deelnemers nog meer meedoen.
- In het onderzoek zijn professionals en cliënten betrokken. Bij het evalueren van de (inter)nationale studies bestaat hun betrokkenheid uit participatie via focusgroepen en/of interviews waarin ze vanuit hun expertise reflecteren op de praktische toepassing van de kernelementen op hun doelgroep, de verhouding van de interventie tot de door hen tot nu toe gehanteerde interventies, de condities voor de implementatie van de nieuwe interventie, en de mogelijke verbeteringen (en eventuele verslechtingen) ten opzichte van de huidige praktijk voor de deelnemers.

3. Sterke aanwijzingen voor effectiviteit

- Zie de algemene criteria voor effectiviteit, plus de criteria voor eerste en goede aanwijzingen. In aanvulling daarop:
- De opzet van het empirisch onderzoek is transdisciplinair van aard en combineert theoretisch en praktisch gefundeerde onderzoeksvragen met een daarbij passende onderzoeksopzet. De opzet maakt duidelijk of het gaat om een planevaluatie, een procesevaluatie of een productevaluatie en hanteert adequate dataverzamelmethode(n). Mogelijke onderzoeksopzetten zijn een (quasi-

)experimentele opzet, herhaalde casestudies, een vergelijkende cohort- of casestudie, realistische evaluatie, een monitor, of een historische veranderingstudie van hoge kwaliteit waarbij uitvoerende professionals als mede-onderzoekers zijn betrokken. De studies zijn in de praktijk uitgevoerd en hebben een follow-up van minstens 6 maanden.

- Het onderzoek heeft naar voren gebracht wat eventuele problemen zijn bij de implementatie en welke variatie in effect bestaat bij verschillende cliënten/doelgroepen, wat effecten zijn van organisatorische en institutionele settings en van verschillende benaderingen op cliënten. Beschreven wordt wat de resultaten kunnen betekenen voor het ontwerp van de interventie.
- Het aantal studies kan sterk uiteenlopen, afhankelijk van de kwaliteit en de aard van de studie. Vuistregels voor het minimum zijn:
 - o Er zijn minstens twee Nederlandse studies naar de voorliggende interventie met een sterke of zeer sterke bewijskracht of één Nederlandse studie naar de voorliggende interventie in combinatie met minstens één (inter-)nationale studie naar deze of een vergelijkbare interventie met sterke of zeer sterke bewijskracht. In het laatste geval is aannemelijk gemaakt dat het doel, de doelgroep, de aanpak en het theoretisch model van de vergelijkbare interventie voldoende van toepassing is op de Nederlandse situatie en/of de voorliggende interventie. De erkenningscommissie oordeelt uiteindelijk over de vergelijkbaarheid.
 - o Bij herhaalde of vergelijkende casestudies zijn er minstens 3 cases uitgevoerd door verschillende behandelaars in verschillende condities of minstens 5 cases uitgevoerd in één setting.

Referenties

- Anderson, M., Cosby, J., Swan, B., Moore, H., & Broekhoven, M. (1999). The use of research in local health service agencies. *Social Science & Medicine* 49(8), 1007-1019.
- Bellamy, J. L., Bledsoe, S. E., & Traube, D. E. (2006). The current state of evidence-based practice in social work: A review of the literature and qualitative analysis of expert interviews. *Journal of Evidence-Based Social Work* 3(1), 23-48.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters. A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology* 41(3-4), 327-350.
- Barkham, M., & Mellor-Clark, J. (2003). Bridging evidence-based practice and practice-based evidence: developing a rigorous and relevant knowledge for the psychological therapies. *Clinical Psychology & Psychotherapy* 10(6), 319-327.
- Cartwright, N. (2007). Are RCTs the gold standard? *BioSocieties* 2(1), 11-20.
- Davies, P. (2005). *The Campbell Collaboration*, paper presented on the 1^e Campbell Seminar Evidence-based practice. Belgian Centre for Evidence-based Medicine, 13 april, België, Leuven.
- Davies, K., & Gray, M. (2016). The place of service-user expertise in evidence-based practice. *Journal of Social Work* 17(91) 3-20.
- Dehue, T. (2002). Over de assumpties van RCT's geïllustreerd aan het Nederlandse experiment met verstrekking van heroïne. *Maandblad voor de Geestelijke Gezondheidszorg* 57, 230-249.
- Drenth, P. (2016). Is psychologie wel een wetenschap? *De Volkskrant*, 12-8-2016, p. 19
- Erkenning van interventies. Criteria voor gezamenlijke kwaliteitsbeoordeling 2015-2018 (2015)*. RIVM, NJI, NISB, Trimbos instituut, Vilans, MOVISIE, NCJ
- Flyvbjerg, B. (2006), Five misunderstandings about case-study, *Qualitative Inquiry* 12 (2), 219-245
- Gibbons, M., Limoges, C., Nowotny, H., Schwartzmann, S., Scott, P., & Trow, M. (1995). *The New Production of Knowledge. The dynamics of Science and Research in Contemporary Societies*. London: Sage Publications.
- Gibbs, L. E., & Gambrill, E. (2002). Evidence-based practice: Counterarguments to objections. *Research on Social Work Practice* 12(3), 452-476.
- Gordon, J., & Cooper, B. (2010). Talking knowledge—practising knowledge: A critical best practice approach to how social workers understand and use knowledge in practice. *Practice: Social Work in Action* 22(4), 245-257.
- Gray, M., Sharland, E., Heinsch, M., & Schubert, L. (2015). Connecting Research to action. Perspectives on research utilisation. *British Journal of Social Work* 45(7) 1952-1967.
- Groot, A.D. de (1961). *Methodologie*. Den Haag.
- Ioannidis, J.P. A. (2005). Why most published research findings are false, *PLoS Med* 2(8): e124. DOI:10.1371/journal.pmed.0020124
- Kirk, S. A., & Reid, W. J. (2002). *Science and Social Work: A Critical Appraisal*. New York: Columbia University Press
- Laan, G. van der (2003). De professional als expert in practice-based evidence. Wat maakt sociale interventies effectief? *Sociale Interventie* 12 (4), 5-16.
- Lub, V., de Groot, N., & Schaafsma, J. (2011). *Polarisatie en radicalisering. De onderbouwing van sociale interventies getoetst*. Utrecht: MOVISIE.
- Lub, V. (2014) *Kwalitatief evalueren in het sociale domein, Meppel: Boom*
- Lub, V. (2017) *Vaste grond. Theorie en theoriegebruik in maatschappelijke activering*. Rotterdam: MOVISIE
- Machielse, A. (forthcoming). Experiential effectivity of social guidance trajectories for socially isolated elderly with complex problems. Findings from an evaluation study using the perceived benefit approach. *Ageing & Society*.

- Major, H.C., & Savin-Baden, M. (2011). Integration of qualitative evidence: towards construction of academic knowledge in social science and professional fields. *Qualitative Research* 11(6) 645–663.
- Martens, R (2016) Zo duivels is het dilemma van praktijkgericht onderzoek niet. *OnderwijsInnovatie* (OI 3).
- McNeece, C. A., & Thyer, B. A. (2004). Evidence-based practice and social work. *Journal of Evidence-Based Social Work* 1(1), 7-25.
- McNeece, C.A., & Thyer, B.A. (2008). Evidence-based practice and social work. *Journal of Evidence-Based Social Work* 1(1), 7-25.
- Michon, H., & de Kroon, H. (2013) En of het werkt? Over de pragmatische RCT. In J. Omlo, M. Bool & P. Rensen (red.) *Weten wat werkt. Passend onderzoek in het sociale domein*. Amsterdam: Uitgeverij SWP.
- Moulaert, F., MacCallum, D., Mehmood A. & Hamdouch, A. (2013). *The International Handbook on Social Innovation*. Cheltenham: Edward Elgar
- Mullen, E.J., & Bacon, W. (2004). A survey of practitioner adoption and implementation of practice guidelines and evidence-based treatments. In A. R. Roberts & K. Yeager (Eds.), *Evidence-Based Practice Manual: Research and Outcome Measures in Health and Human Services*. New York, NY: Oxford University Press.
- OCW (2014) Wetenschapsvisie 2025. Den Haag, Ministerie van Onderwijs, Cultuur en Wetenschappen.
- Omlo, J. (2011). Waar zijn we nu eigenlijk mee bezig? Kwalitatief onderzoek maakt sociale interventies beter. *Tijdschrift voor Sociale Vraagstukken* 65(10), 26-29.
- Owen, J. (2007). *Program evaluation. Forms and approaches*. New York: The Guilford Press.
- Pawson R. & N. Tilley (1997) *Realistic Evaluation*, Londen: Sage.
- Potting, M., Sniekers, M., Lamers, C., & Reverda, N. (2010). Legitimizing social work: The practice of reflective professionals. *Journal of Social Intervention: Theory and Practice* 19(3), 6-20.
- Ragin, Ch. (2000) *Fuzzy-Set Social Science*, Chicago: University of Chicago Press
- Sackett, D., Straus, S. Richardson, W., Rosenberg, W., & Haynes, R. (2000). *Evidence Based Medicine. How to Practice and Teach EBM*. Edinburgh: Churchill Livingstone.
- Scholte, M. & A. Sprinkhuizen (2012). *De sociale kwestie hervat. De Wmo en sociaal werk in transitie*. Houten: Bohn Stafleu, Van Loghum.
- Scriven, M. (2008). A summative evaluation of RCT methodology & an alternative approach to causal research. *Journal of Multidisciplinary Evaluation* 5(9), 11-24.
- Soydan, H. & Palinkas, L.A. (2016). *Evidence-Based Practice in Social Work. Development of a New Professional Culture*. New York; Routledge.
- Steyaert, J., Biggelaar, T. van, & Peels, J. (2010). *De bijziendheid van evidence based practice: Beroepsinnovatie in de sociale sector*. Amsterdam: SWP.
- Swanborn, P. *Evalueren*. Meppel: Boom.
- Tavecchio, L., & Gerrebrands, M. (2012). *Bewijsvoering binnen praktijkgericht onderzoek. Methodologische en wetenschapstheoretische reflecties op de onderbouwing van professionele interventies*. Den Haag: Boom Lemma.
- Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. Thousand Oaks, CA: SAGE.
- Veerman, J. W., & van Yperen, T. A. (2007). Degrees of freedom and degrees of certainty: A developmental model for the establishment of evidence-based youth care. *Evaluation and Program Planning* 30(2), 212-221.
- Vlaeminck, H. de (2005). *Het gebruik van casuïstiek in het sociaal werk*. Gent: Academia Press.
- Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., & Saul, J. (2008). Bridging the gap between prevention research and practice: The interactive systems framework for dissemination and implementation. *American Journal of Community Psychology* 41(3-4), 171-181.

Zwet, R. van der, Beneken genaamd Kolmer, D., & Schalk, E. (2011). Op weg naar een interactieve benadering van evidence-based werken in de sociale sector in Nederland. *Journal of Social Intervention: Theory and Practice* 20(4): 62-78.

Zwet, R. van der, Booijnk, M., & Kok, E. (2015). *Een kwalitatief onderzoek naar het gebruik van effectieve sociale interventies. Bevorderende en belemmerende factoren*. Utrecht: Movisie.

Criteria voor de (causale) bewijskracht van empirische studies

Bewijskracht design	Kenmerken studie
Zeer sterk	Hier gelden dezelfde criteria als bij het onderliggende niveau, met als verschil: <ul style="list-style-type: none"> Er is een experimentele onderzoeksopzet (d.w.z. er is een aselecte toewijzing van subjecten aan onderzoeksgroepen) of er is een andere opzet die de causale relatie tussen interventie en effect voldoende overtuigend aantoont.
Sterk	Hier gelden dezelfde criteria als bij het onderliggende niveau, met als toevoeging: <ul style="list-style-type: none"> Er is een follow-up (vuistregel: 6 maanden) of er is een andere opzet die voldoende zicht biedt op de stabiliteit van de resultaten.
Vrij sterk	Hier gelden dezelfde criteria als bij het onderliggende niveau, met als verschil: <ul style="list-style-type: none"> De studie is in de praktijk uitgevoerd / is representatief voor de praktijk.
Redelijk	Hier gelden dezelfde criteria als bij de onderliggende niveaus, met als verschil: <ul style="list-style-type: none"> Het is een onderzoek met een (quasi-) experimentele en een controle-groep (care as usual) of een herhaald N=1 onderzoek met een baseline of een time series design met een (multiple) baseline of alternating treatments of een studie naar de samenhang tussen de mate waarin een interventie is toegepast en de mate waarin bedoelde uitkomsten zijn opgetreden. Het design is van goede kwaliteit. Het onderzoek is niet in de praktijk is uitgevoerd / is niet representatief voor de praktijk of de representativiteit voor de praktijk is onbekend.
Matig	Hier gelden dezelfde criteria als bij het onderliggende niveau, met als toevoeging: <ul style="list-style-type: none"> De resultaten zijn vergeleken met ander onderzoek naar de effecten van de gebruikelijke situatie, handelwijze of zorg (care as usual) of een andere zorgvorm bij een soortgelijke doelgroep.
Licht	De eisen die op dit niveau gelden zijn: <ul style="list-style-type: none"> Het gemeten effect heeft betrekking op het doel en de doelgroep van de interventie. De meting is verricht met betrouwbare en valide onderzoeksmethoden en/ of instrumenten. Er zijn resultaten beschikbaar voorafgaand aan, bij de start of tijdens de interventie en bij einde van de interventie (zgn. voor- en nameting). Bij onderzoek naar de ervaren effectiviteit volstaan de resultaten van een nameting Het onderzoek is zodanig gedocumenteerd, dat replicatie van de uitvoering van de studie mogelijk is. Voor kwantitatief onderzoek geldt: de resultaten zijn met een adequate statistische techniek geanalyseerd, op significantie getoetst en er is een geaccepteerde uitkomstmaat (zoals Cohen's d of een Odds Ratio) berekend of te berekenen.
Zwak	De studie voldoet niet aan de eisen die minstens aan een empirisch onderzoek te stellen zijn.