

Checklists kwalitatief evaluatieonderzoek en mixed methods onderzoek

Met toelichting



Colofon

Dit is een herziene versie van de publicatie 'Kwalitatief evaluatieonderzoek op waarde schatten. Een checklist met passende criteria' van het Bureau voor Sociale Argumentatie/Vasco Lub (2020).

In opdracht van

Kenniscentrum Sport & Bewegen, Movisie, Nederlands Jeugdinstituut, RIVM, Trimbos-instituut en Vilans, verenigd in het Samenwerkingsverband Erkenning van Interventies.

2

Contactpersoon:

Peter Rensen (Movisie)

E-mail: p.rensen@movisie.nl

Auteurs

Vasco Lub (Bureau voor Sociale Argumentatie)

Peter Rensen (Movisie)

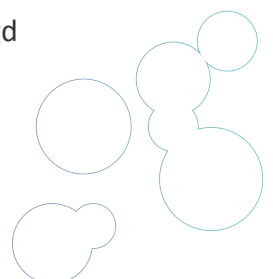
Datum

Augustus 2022

Met subsidie van

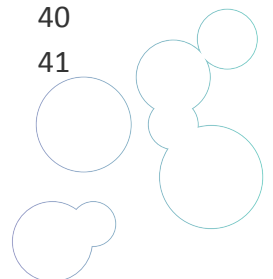
Ministerie van Volksgezondheid, Welzijn en Sport

© Met duidelijke bronvermelding mag alles uit deze uitgave worden vermenigvuldigd en/of openbaar gemaakt door middel van sociale media, druk, fotokopie, of op enige andere wijze.



Inhoudsopgave

1.	Inleiding	4
1.1	Vooraf	4
1.2	Gebruik van de checklists bij erkenning	5
2.	Checklist kwalitatief evaluatieonderzoek	6
3.	Checklist mixed methods onderzoek	10
4.	Toelichting op de beoordelingssystematiek	11
4.1	Drempelcriteria	11
4.2	Wijze van beoordeling	11
4.3	Puntentelling	12
5.	Toelichting op de criteria uit de checklist kwalitatief evaluatieonderzoek	13
5.1	Drempelcriteria	13
5.2	Context en achtergrond	14
5.3	Methodologie en data	16
5.4	Analyse en verslaglegging	22
5.5	Bijdrage en inbedding kennis	27
6.	Toelichting checklist mixed methods onderzoek	29
6.1	Vertrekpunten onderzoek	29
6.2	Resultaatanalyse	30
6.3	Conclusies	31
7.	Bibliografie	33
8.	Verantwoording	34
8.1	Aanleiding en opdracht	34
8.2	Ontwikkeling kwalitatief evaluatieonderzoek	34
8.3	Constructie checklists	35
	Bijlagen	39
	Bijlage 1 Verzamelde (internationale) checklists voor kwalitatief onderzoek	39
	Bijlage 2 Geraadpleegde deskundigen	40
	Bijlage 3 Klankbordcommissie	40
	Bijlage 4 Geraadpleegde deskundigen mbt. toepassing erkenningstraject	41



1. Inleiding

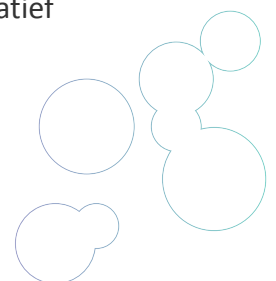
1.1 Vooraf

(Beleids)interventies worden steeds vaker kwalitatief getoetst. Het gaat niet alleen om cijfers maar ook om de mensen achter de cijfers: de cliënten, burgers, professionals en beleidsmakers. Hoe vatten zij deze interventies op? Kwalitatief evaluatieonderzoek wint aan belangstelling. Maar wat is *goed* kwalitatief evaluatieonderzoek? Op basis van welke criteria kunnen we dit onderzoek beoordelen? Deze vraag kwam op in het erkenningstraject waar zes kennisinstituten samenwerken bij het beoordelen van interventies op het terrein van bewegen & sport, gezondheidsbevordering, jeugdgezondheidszorg, jeugdzorg/brede jeugdhulp, ontwikkelingsstimulering, langdurende geestelijke gezondheidszorg, ouderenzorg, gehandicaptenzorg en het sociaal domein. Deze vraag heeft geleid tot twee checklists met toetsingscriteria: Checklist kwalitatief evaluatieonderzoek en Checklist mixed methods onderzoek.

Aan de samenstelling van de checklists is uitgebreid vooronderzoek voorafgegaan (zie: Verantwoording). Dit heeft geleid tot de publicatie: 'Kwalitatief evaluatieonderzoek op waarde schatten. Een checklist met passende criteria' van het Bureau voor Sociale Argumentatie/ Vasco Lub (2020). Vervolgens zijn de checklists toepasbaar gemaakt voor het erkenningstraject. Hieronder vindt u aanwijzingen voor gebruik van de checklists in het erkenningstraject. Natuurlijk kunnen de checklists ook buiten het erkenningstraject worden benut.

Voor de duidelijkheid: *kwalitatief* onderzoek omvat al het onderzoek dat variabelen gebruikt die niet gemeten zijn op interval- of rationiveau (Swanborn, 2004). Bekende methodes zijn onder meer: interviews, focusgroepen, observaties en documentanalyse. Kwalitatief *evaluatie*onderzoek richt zich specifiek op inschattingen over de werkzaamheid van een bepaalde methodiek, werkwijze of (gedrags)interventie. Hiervoor gelden specifieke methodologische en analytische mogelijkheden.

De checklist kwalitatief evaluatieonderzoek betreft algemene criteria voor kwalitatief evaluatieonderzoek. De criteria geven aan in hoeverre een onderzoek betrouwbaar en valide is. De checklist mixed methods evaluatieonderzoek betreft richtlijnen om vast te stellen of gecombineerd kwalitatief en kwantitatief evaluatieonderzoek valide en betrouwbaar is. Daarbij ligt de nadruk op de meerwaarde van de combinatie van kwalitatief en kwantitatief evaluatieonderzoek. Het kwalitatieve onderdeel van mixed methods evaluatieonderzoek kan ook worden bekeken met de checklist kwalitatief evaluatieonderzoek. Voor kwantitatief onderzoek zijn al criteria beschikbaar.



1.2 Gebruik van de checklists bij erkenning

Leidraad en scorelijst

Met de checklist kwalitatief evaluatieonderzoek is gestreefd naar het beoordelen op basis van de kracht van kwalitatieve evaluatietechnieken en -methoden. Hiermee wordt een objectieve maatstaf ingebouwd voor de beoordeling. De ene methode is nu eenmaal krachtiger dan de andere. De checklist biedt hierbij houvast. Voor de checklist mixed methods onderzoek geldt hetzelfde maar dan met betrekking tot evaluatiemethoden en -technieken van mixed methods onderzoek.

Beoordeling onderzoek is nog geen beoordeling interventie

De checklists zijn bedoeld om het *onderzoek* te beoordelen. De checklists zijn niet bedoeld om te beoordelen of de *interventie* waar het onderzoek over gaat effectief is. Dat wordt vastgesteld aan de hand van de criteria van het erkenningstraject¹.

Wel stellen de checklists vereisten aan het evaluatief gehalte van het onderzoek.

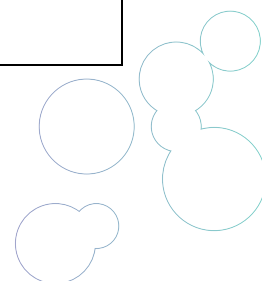
Verschillende criteria in de checklists hebben hier betrekking op. Zo wordt er gekeken naar: 'De onderzoeksvragen zijn geformuleerd met het oog op de beoordeling van effectiviteit (koppeling interventiedoelen)'. Doel van deze vragen is echter het toetsen van het onderzoek en niet het toetsen van de effectiviteit van de interventie. Wanneer het onderzoek 'Goed' wordt bevonden, speelt het wel een rol bij het toetsen van de effectiviteit van de interventie. Maar dat vraagstuk wordt door de commissieleden gezien als onderdeel van al het beschikbare (onderzoeks)materiaal.

¹ De bewijskrachttabel die in het erkenningstraject gebruikt wordt geeft aan aan welke vereisten een onderzoek moet voldoen om mee te tellen voor een bepaald niveau van erkenning. Aan de hand van de checklists kan worden vastgesteld aan welke kwalificatie van de bewijskrachttabel het onderzoek voldoet. En voor welk niveau van erkenning het onderzoek ondersteuning biedt. Welk soort onderzoek vereist is om aan een bepaald niveau van erkenning te voldoen is weergegeven in de criteria voor erkenning. De bewijskrachttabel en de criteria voor erkenning zijn te vinden in de algemene informatie op de websites van de databanken.

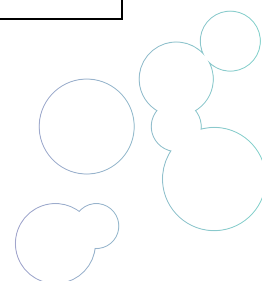


2. Checklist kwalitatief evaluatieonderzoek

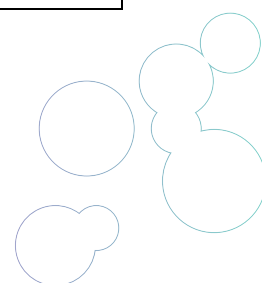
Checklist kwalitatief evaluatieonderzoek	Goed	Twijfel	Onvoldoende	N.v.t.
Drempelcriteria				
Het onderzoek is niet in strijd met de Nederlandse gedragscode integriteit				
In de evaluatie is een centrale onderzoeksvraag of -vragen geformuleerd				
De onderzoeksvragen zijn geformuleerd met het oog op of ruimte voor de beoordeling van effectiviteit (koppeling interventiedoelen)				
Het onderzoek bevat een toelichting op de gebruikte methoden				
Het onderzoek bevat een beschrijving van het analysekader of de wijze van data-analyse				
Het respondentensample is toegelicht in het analysekader				
Onderzoek door / niet door				
1. Context en achtergrond				
Onderzoeksopzet en vragen				
1. De rapportage bevat een beschrijving van de onderzoeksopzet / het evaluatieontwerp				
2. De rapportage bevat een beschrijving van de geëvalueerde interventie				
3. De rapportage bevat een beschrijving van de interventiedoelen en het gewenste resultaat				
Locaties en settings				
4. De rapportage bevat een beschrijving van de onderzoekslocatie(s) en setting(s) van de evaluatie				
5. De keuze voor de specifieke onderzoekslocatie(s) en setting(s) ligt voor de hand gegeven de onderzoeksvragen				
6. De positie en de rol van de onderzoekers tijdens de evaluatie zijn geëxpliciteerd				
7. De rapportage beschrijft de maatschappelijke context van de interventie en evaluatie				
Toelichting				



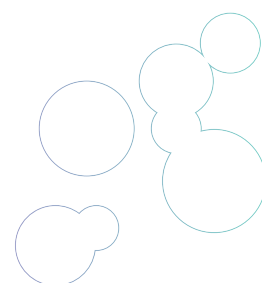
Checklist kwalitatief evaluatieonderzoek	Goed	Twijfel	Onvoldoende	N.v.t.
2. Methodologie en data				
Algemene methoden dataverzameling				
8. Er is sprake geweest van langdurige betrokkenheid van de onderzoekers in het veld ('prolonged engagement')				
9. Het onderzoek kenmerkt zich door verschillende evaluatiemomenten ('repeated measures')				
10. In het onderzoek zijn via multi-site studies controlesituaties gecreëerd om licht te werpen op causale mechanismen ('counterfactual condition')				
11. Het onderzoek belicht perspectieven van verschillende groepen die relevant zijn voor de evaluatie ('fair dealing')				
12. Onderzoeksmethoden richten zich op de facilitators en barriers bij respondenten				
Observatie				
13. De evaluatie maakt gebruik van observaties				
14. Observaties zijn gebruikt om voorwaarden van effectiviteit te toetsen ('necessary conditions')				
15. Observaties zijn gebruikt om veronderstelde effecten van de interventie te toetsen				
16. Observaties zijn gebruikt om licht te werpen op causale mechanismen van de interventie (oorzaak- en gevolgrelaties werkzame elementen)				
17. Interviews en andere gespreksmethoden hadden zoveel mogelijk plaats na gedane observaties (verzamelen van accounts op basis van action)				
Respondentensample				
18. Respondenten zijn geselecteerd op basis van relevante kenmerken van de bredere doelgroep(en) van de evaluatie ('purposive sampling')				
19. De omvang van het respondentensample en de data die hiermee zijn gegenereerd, zijn afdoende gegeven de interventiekenmerken en het evaluatiedoel				
20. De rapportage bevat een toelichting op eventueel 'gemiste' respondenten (non-participatiedoelgroepen)				
21. De onderzoekers hebben gereflecteerd op hoe respondentengedrag of -zinswijzen tijdens de evaluatie door hen mogelijk zijn beïnvloed				
Toelichting				



Checklist kwalitatief evaluatieonderzoek	Goed	Twijfel	Onvoldoende	N.v.t.
3. Analyse en verslaglegging				
Illustraties resultaten				
22. Het verslag bevat illustraties van de originele data (bijv. citaten), zodanig dat de lezer een beeld krijgt van de evaluatie en de uitkomsten				
23. De onderzoeksresultaten zijn gedetailleerd beschreven aan de hand van citaten, fragmenten en contextbeschrijving ('thick description')				
Analysetechnieken en procedures				
24. Het verslag bevat een toelichting wanneer een punt van data-saturatie is bereikt en hoe dit is vastgesteld				
25. De analyse/codering van de data is gedaan door verschillende onderzoekers (peer analyse / onafhankelijke codering)				
26. Resultaten van het onderzoek zijn gecontrasteerd met andere databronnen ('triangulatie')				
27. Interview- of observatieverslagen of resultaten zijn ter controle teruggekoppeld aan respondenten ('member check')				
28. Onderzoekers hebben het onderzoeksproces en de keuzes daarin nauwgezet en chronologisch gedocumenteerd ('audit trail')				
29. In de data-analyse is bewust naar deviante 'uitschieters' gezocht en deze worden verklaard ('negative case selection')				
Constructie conclusies en verklaring modellen				
30. Op basis van de bevindingen is een verklarend model geconstrueerd voor de (veronderstelde) causale relaties van de interventie				
31. Op basis van de bevindingen wordt een analyse gepresenteerd die iets zegt over de effectiviteit van de interventie				
32. De evaluatieresultaten worden besproken in het licht van de onderzoeksvragen (samenhang resultaten en analyse)				
33. Conclusies van de evaluatie zijn coherent en geloofwaardig ('credibility')				
34. Conclusies van de evaluatie zijn voorgelegd aan externe deskundigen ('peer debriefing')				
35. De onderzoekers gaan in op hun eigen meningen en belangen die van invloed kunnen zijn op de interpretatie van de resultaten (reflexiviteit)				
Toelichting				

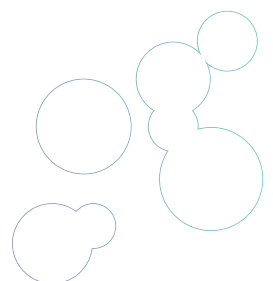


Checklist kwalitatief evaluatieonderzoek	Goed	Twijfel	Onvoldoende	N.v.t.
4. Bijdrage en inbedding kennis				
Koppeling literatuur en theorie				
36. De vertrekpunten van de evaluatie zijn ingebed in bestaande theorievorming				
37. De resultaten van de evaluatie worden in een bredere context geplaatst door aan te sluiten bij eerder relevant onderzoek/literatuur/theorie				
Kennisontwikkeling en implicaties van de evaluatie				
38. De onderzoekers bespreken expliciet welke nieuwe inzichten de evaluatie heeft opgeleverd				
39. De onderzoekers bespreken aan de hand van de resultaten wat er in de toekomst nodig is aan kennisontwikkeling over de interventie (lacunes)				
40. De onderzoekers bespreken of en hoe de resultaten kunnen worden gegeneraliseerd naar andere interventiecontexten				
41. De onderzoekers bespreken aan de hand van de resultaten beleids- en praktijkimplicaties				
Toelichting				
Totale score				



3. Checklist mixed methods onderzoek

Checklist mixed methods onderzoek	Goed	Twijfel	Onvoldoende	N.v.t.
Vertrekpunten onderzoek				
1. De onderzoekers onderbouwen de keuze voor mixed methods				
2. De kwantitatieve en kwalitatieve onderzoeksmethoden zijn ingezet om gerelateerde aspecten van dezelfde onderzoeksvraag te beantwoorden				
3. Het onderzoek streeft naar een integratie van kwantitatieve en kwalitatieve onderzoeksresultaten en analysetechnieken				
4. Het onderzoek bevat een beschrijving van de analyse-strategie die is gebruikt om deze integratie te bewerkstelligen				
Resultaatanalyse				
5. De resultaten van de kwalitatieve en kwantitatieve onderzoekslijnen zijn in samenhang geanalyseerd, gecontrasteerd en gerapporteerd ('triangulatie')				
6. Resultaten van beide onderzoekslijnen krijgen in de rapportage evenveel prioriteit				
7. Het onderzoek expliciteert wanneer en hoe bepaalde uitkomsten van de verschillende onderzoekslijnen zijn gebruikt en hoe zij zich tot elkaar verhouden				
8. Het onderzoek expliciteert welke onderzoekers verantwoordelijk waren voor welke methoden en analyses en hoe zij met elkaar afstemden				
Conclusies				
9. Het onderzoek levert verklarende informatie op over de geëvalueerde interventie				
10. Conclusies die voortkomen uit beide onderzoekslijnen worden aangewend om metaconclusies te trekken				
11. De onderzoekers bespreken welke nieuwe inzichten de mixed methods benadering heeft opgeleverd				
Toelichting				



4. Toelichting op de beoordelingssystematiek

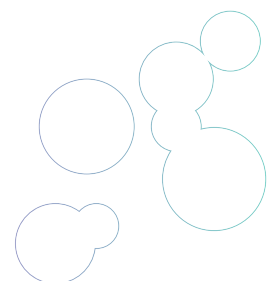
4.1 Drempelcriteria

Als voorportaal van de checklist kwalitatief evaluatieonderzoek worden enkele drempelcriteria getoetst (zie paragraaf 5.1). Aan de hand van deze drempelcriteria wordt bepaald of het aangedragen onderzoek überhaupt bruikbaar is qua integriteit, niveau en inhoud voor de erkenning van een interventie. Komt het onderzoek erdoor, dan wordt het vervolgens beoordeeld op criteria die zijn gerangschikt naar vier hoofdthema's: (1) Context en achtergrond, (2) Methodologie en data, (3) Analyse en verslaglegging en (4) Bijdrage en inbedding kennis. Deze worden vanaf paragraaf 5.2 verder toegelicht.

Voor de checklist mixed methods onderzoek is afgezien van drempelcriteria. De drempelcriteria voor kwantitatief evaluatieonderzoek zijn elders beschikbaar en de drempelcriteria voor kwalitatief evaluatieonderzoek zijn al te vinden in de checklist kwalitatief evaluatieonderzoek.

4.2 Wijze van beoordeling

De beoordeling van een onderzoek gebeurt aan de hand van stellingen met criteria die op een vierpuntsschaal worden beoordeeld. Met de stellingen wordt een standaard vastgesteld voor wat sterk kwalitatief evaluatieonderzoek is. Een voorbeeld op het gebied van methodologie en data: het format vraagt niet: 'Is de wijze van respondentselectie duidelijk? (*convenience*, *purposive*, sneeuwbal, enz.)' maar stelt: 'De wijze van respondentselectie had plaats op basis van *purposive sampling*'. Purposive sampling is methodologisch namelijk sterker dan bijvoorbeeld de sneeuwbal-methode of *convenience sampling*. Al naar gelang de mate waarin aan een criterium wordt voldaan, krijgt het onderzoek op dat aspect de beoordeling 'goed', 'twijfel', 'onvoldoende' of 'niet van toepassing'. In de toelichting bij de verschillende onderdelen worden aan de hand van fictieve situaties enkele voorbeelden gegeven van hoe te beoordelen. Als uit de rapportage niet is op te maken of aan een bepaald criterium is voldaan, scoort het op dat punt in principe 'onvoldoende'. De beoordeling 'niet van toepassing' wordt enkel gegeven als om praktische of inhoudelijke redenen het criterium niet kon worden toegepast.



4.3 Puntentelling

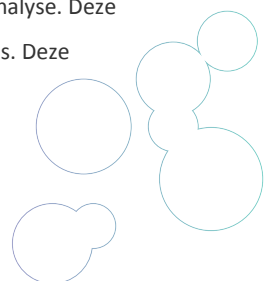
De checklist kwalitatief evaluatieonderzoek kent een puntentelling. Alle afzonderlijke criteria in de checklist wegen even zwaar². Voor elke 'goed' krijgt het onderzoek één punt. Het beoordeelde onderzoek dient uiteindelijk een minimum aantal 'goed' te halen om als 'sterk' te worden beoordeeld. Wanneer het onderzoek op meer dan de helft van het aantal punten 'goed' scoort wordt het als 'sterk' beoordeeld. Voor kwalitatief evaluatieonderzoek ligt de drempelwaarde daarmee op 21 punten. Dit is gebaseerd op een analyse van de haalbaarheid van de criteria. Van de gehele set van 41 criteria zijn 21 criteria gebruikelijke onderdelen van kwalitatief evaluatieonderzoek. Van deze set criteria kan verwacht worden dat onderzoekers er aandacht aan hebben besteed.

De eindscore zegt dus iets over de kwaliteit van de evaluatie en daarmee over de mate waarin uitspraken zijn te doen over de werkzaamheid van de onderzochte interventie. Voor een kwalitatief evaluatieonderzoek kan zodoende het oordeel zijn dat het onderzoek sterke uitspraken toelaat over de effectiviteit van de interventie (score 21 en hoger) of slechts beperkte uitspraken of geen uitspraken hierover toelaat (score 20 en lager).

Doordat alleen een punt wordt verkregen bij 'goed', is de checklist feitelijk binair. Echter, de categorieën 'twijfel' en 'niet van toepassing' bieden de mogelijkheid rekening te houden met de specifieke omstandigheden van een evaluatie en geven ruimte voor nuance. Deze categorieën leveren geen punten op, maar de beoordelaars kunnen de waardering op deze categorieën eventueel wel laten meewegen in hun keuze om het onderzoek alsnog te accepteren, mocht het onderzoek (net) onder de 20 scoren. Het is in dit verband bovendien belangrijk om nogmaals te vermelden dat de eindscore enkel een beoordeling omvat van de sterkte van het desbetreffende onderzoek (validiteit en betrouwbaarheid). Het is dus niet een eindoordeel over de werkzaamheid van een interventie. Die beoordeling heeft plaats op basis van meer bronnen en inzichten dan één kwalitatief onderzoek alleen.

Anders dan de checklist kwalitatief evaluatieonderzoek, kent de checklist mixed methods onderzoek geen puntentelling.

² De checklist hanteert een weging omdat het merendeel van de criteria betrekking heeft op methodologie en analyse. Deze thema's hebben dus relatief meer gewicht dan de thema's context & achtergrond en bijdrage & inbedding kennis. Deze verdeling ligt voor de hand aangezien methodologische - en analytische criteria het meest direct van toepassing zijn op de betrouwbaarheid en validiteit van een onderzoek.



5. Toelichting op de criteria uit de checklist kwalitatief evaluatieonderzoek

In dit deel wordt een toelichting gegeven op de hoofd- en subthema's van de checklist en de afzonderlijke criteria. De toelichting volgt de genummerde rubricering van de checklist. Voor elk thema en criterium worden de centrale begrippen gedefinieerd en toegelicht en kun relevantie voor de beoordeling van kwalitatief evaluatieonderzoek beargumenteerd. Daarnaast wordt een illustratie gegeven van hoe de criteria te beoordelen (aan de hand van fictieve praktijkvoorbeelden).

5.1 Drempelcriteria

- Het onderzoek is niet in strijd met de Nederlandse gedragscode wetenschappelijke integriteit³.
- In de evaluatie is een centrale onderzoeksvraag of -vragen geformuleerd.
- De onderzoeksvragen zijn geformuleerd met het oog op of ruimte voor de beoordeling van effectiviteit (koppeling interventiedoelen).
- Het onderzoek bevat een toelichting op de gebruikte methoden.
- Het onderzoek bevat een beschrijving van het analysekader of de wijze van data-analyse.
- Het respondentensample is toegelicht in het analysekader.

Elke beoordeling van een evaluatie start met het toetsen van drempelcriteria. Deze criteria moeten 'goed' of 'twijfel' scoren wil de evaluatie in aanmerking komen voor de feitelijke beoordeling (zij mogen dus niet 'onvoldoende' scoren). Deze criteria tellen niet mee in de puntentelling. Het onderzoek mag niet in strijd zijn met het normatieve kader dat door de wetenschapsgemeenschap gezamenlijk is vastgesteld. Er moet een centrale onderzoeksvraag of -vragen zijn geformuleerd en de onderzoeksvragen moeten zijn geformuleerd met het oog op de beoordeling van effectiviteit. Dit houdt in dat het doel, of de doelen, van de interventie wordt geëxpliciteerd. Dit criterium is relevant omdat een evaluatie zich specifiek richt op een inschatting van de werkzaamheid van een interventie. Het moet dus duidelijk zijn bij welk resultaat de interventie als effectief geldt. Een evaluatie van een interventie die bijvoorbeeld delinquente jongeren voorbereidt op terugkeer in de maatschappij, moet de onderzoeksvraag dus formuleren in termen van die doelstelling. Anders kan niet worden

³ Deze gedragscode is in 2018 onderschreven door: KNAW (Koninklijke Nederlandse Academie van Wetenschappen), NFU (Nederlandse Federatie van Universitair Medisch Centra), NWO (Nederlandse Organisatie voor Wetenschappelijk Onderzoek), TO2-federatie (samenwerkende Toegepast Onderzoek Organisaties, VSNU (Vereniging van Universiteiten) en de Vereniging Hogescholen. Het betreft een normatief kader voor methodologisch goed en ethisch onderzoek.



beoordeeld in hoeverre de interventie heeft bijgedragen aan een succesvolle maatschappelijke re-integratie van de jongeren.

Voorts moet het onderzoek een toelichting bevatten van de gebruikte methoden en de wijze van data-analyse of het analysekader beschrijven. Met een toelichting op de gebruikte methoden en de wijze van data-analyse expliciteren de onderzoekers hoe ze de onderzoeksvraag hebben beantwoord. In het analysekader moet het respondentensample zijn toegelicht. De methoden en het analysekader moeten dus omschreven zijn, omdat anders niet is vast te stellen hoe valide of betrouwbaar de bevindingen zijn en verdere beoordeling geen zin heeft.

5.2 Context en achtergrond

Onderzoeksopzet en vragen

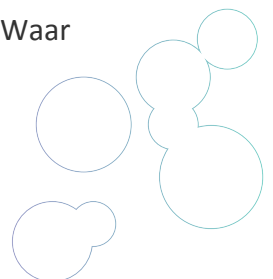
1. De rapportage bevat een beschrijving van de onderzoeksopzet/het evaluatieontwerp.
2. De rapportage bevat een beschrijving van de geëvalueerde interventie.
3. De rapportage bevat een beschrijving van de interventiedoelen en het gewenste resultaat.

Het onderdeel Context en achtergrond toetst zeven criteria verdeeld naar twee subthema's. Het eerste subthema gaat over de onderzoeksopzet en -vragen. Beoordeeld wordt of de rapportage een beschrijving van de onderzoeksopzet/het evaluatieontwerp bevat (1), of de geëvalueerde interventie is beschreven (2) met een explicitering van de interventiedoelen en het gewenste resultaat (3).

Locaties en settings

4. De rapportage bevat een beschrijving van de onderzoekslocatie(s) en setting(s) van de evaluatie.
5. De keuze voor de specifieke onderzoekslocatie(s) en setting(s) ligt voor de hand gegeven de onderzoeksvragen.
6. De positie en de rol van de onderzoekers tijdens de evaluatie zijn geëxpliciteerd.
7. De rapportage beschrijft de maatschappelijke context van de interventie en evaluatie.

Het tweede subthema gaat over de locaties en settings van het onderzoek. De rapportage moet hiervan een beschrijving geven (4). Een onderzoekslocatie heeft meestal een geografische afbakening (bijvoorbeeld een wijk of een gebouw), de setting gaat over een omschrijving van de omgevingscontext binnen die locatie. Ook die moet helder zijn. Waar binnen de wijk is er bijvoorbeeld geobserveerd en op welke tijdstippen? Hadden interviews plaats in een openbaar gebouw of bij respondenten thuis? Enzovoort.

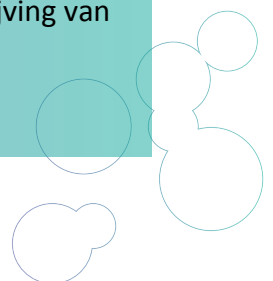


De keuze voor die locaties en settings moet bovendien voor de hand liggen, gegeven de onderzoeksvragen (5). Een evaluatie die bijvoorbeeld wil onderzoeken in hoeverre sportactiviteiten in de buurt contact tussen verschillende etnische bevolkingsgroepen bevorderden, zal meer voor de hand liggen in een multiculturele stadsbuurt dan in een relatief etnisch homogene dorpswijk. De positie en de rol van de onderzoekers tijdens de evaluatie dient hierbij te worden geëxpliciteerd (6). Hebben de auteurs van het evaluatierapport ook het onderzoek uitgevoerd? Wie observeerde of nam interviews af? Is er werk uitbesteed aan externe partijen of onderzoeksassistenten? Enzovoort.

Tot slot is het belangrijk dat de rapportage de maatschappelijke context van de interventie en evaluatie beschrijft (7). Geen enkele sociale of gedragsinterventie speelt zich af in een maatschappelijk vacuüm. Een nauwgezette beschrijving van de maatschappelijke context maakt het mogelijk dat de uiteindelijke geloofwaardigheid van de conclusies van de evaluatie beter op waarde kunnen worden geschat (zie criterium 'Credibility' bij Analyse en verslaglegging). De onderzoekers kunnen bijvoorbeeld aandacht besteden aan de maatschappelijke systemen waarin de interventie zich beweegt en de (verwachte) gedragsverandering van mensen hierin, of een beschrijving geven van de aard van het maatschappelijke probleem waar de interventie voor bedoeld is. Ook moet duidelijk zijn wat de aanleiding was voor de evaluatie en welke partij hiervoor het initiatief nam.

Hoe Context en achtergrond te beoordelen? Enkele fictieve voorbeelden

- ✓ In een evaluatie van een training voor tienermoeders wordt de interventie beschreven, maar wordt dit afgedaan met een korte paragraaf van enkele zinnen. Er blijft veel onduidelijk over hoe de interventie wordt uitgevoerd, wie er bij zijn betrokken en met welk doel deze is ingezet. De evaluatie scoort dan 'twijfel' op het criterium 'De rapportage bevat een beschrijving van de geëvalueerde interventie' (2).
- ✓ In een evaluatie van een interventie gericht op het bevorderen van gezond eetgedrag van ouderen, is de beoogde gedragsverandering bij de doelgroep en hoe de interventie hieraan poogt bij te dragen, uitvoerig toegelicht. De evaluatie scoort dan 'goed' op het criterium 'De rapportage bevat een beschrijving van het interventiedoel of de interventiedoelen' (3).
- ✓ In een evaluatie van een opvoedingscursus voor kwetsbare ouders worden de respondenten en hun kenmerken beschreven, maar wordt nagelaten de context van de interviews te vermelden, waar deze zijn afgenomen, hoe de afspraken hierover met de ouders tot stand kwamen en wie de interviews bij hen afnam. De evaluatie scoort dan 'onvoldoende' op het criterium 'De rapportage bevat een beschrijving van de onderzoekslocatie(s) en setting(s) van de evaluatie' (4).



5.3 Methodologie en data

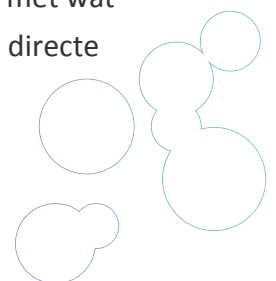
Het onderdeel Methodologie en data toetst de methodologische toepassingen van het onderzoek en de kwaliteit van de data. Hierbij gaat het over de verantwoording van de gebruikte methoden, de wijze van respondentenselectie en de kenmerken daarvan. Bij dit thema wordt beoordeeld hoe robuust die methoden waren en in hoeverre zij licht werpen op evaluatie-inzichten over wat werkt. Ook hier worden specifieke ijkpunten vastgesteld. Bepaalde kwalitatieve methoden geven namelijk een betrouwbaarder beeld van de werkzaamheid van een interventie dan andere methodologische toepassingen. Het gaat om algemene methoden zoals langdurige betrokkenheid bij het veld. Maar ook om specifieke methoden. Observatiemethoden hebben bijvoorbeeld het voordeel dat verschijnselen en menselijk gedrag uit de eerste hand worden waargenomen. Het risico op narratieve misleiding en illusoire causale verbanden wordt daarmee kleiner. De kracht van observatie is dat op een directe manier een beeld kan worden verkregen van hoe een interventie, programma of methodiek feitelijk gestalte krijgt. Via observatie kan worden getoetst of veronderstelde causale mechanismen en effecten ook werkelijk optreden.

Algemene methoden dataverzameling

8. Er is sprake geweest van langdurige betrokkenheid van de onderzoekers in het veld ('prolonged engagement').
9. Het onderzoek kenmerkt zich door verschillende evaluatiemomenten ('repeated measures').
10. In het onderzoek zijn via multi-site studies controlesituaties gecreëerd om licht te werpen op causale mechanismen ('counterfactual condition').
11. Het onderzoek belicht perspectieven van verschillende groepen die relevant zijn voor de evaluatie ('fair dealing').
12. Onderzoeksmethoden richten zich op de facilitators en barriers bij respondenten.

De wijze waarop data worden verzameld kan sterk bijdragen aan de kracht van het onderzoek. Het gaat hierbij om algemene aspecten die elke onderzoeksmethode kunnen versterken.

Het (diepte)interview is waarschijnlijk de bekendste kwalitatieve dataverzamelmethode. Reflecties van respondenten op eigen gedrag, kennis en getuigenissen zijn echter gevoelig voor bias en andere narratieve en psychologische misleidingen, vooral in de context van evaluatievragen. Antwoorden op interviewvragen corresponderen vaak niet met uitkomsten van observationele studie: wat mensen zeggen dat ze doen komt niet altijd overeen met wat zij daadwerkelijk doen. Als het gaat om het genereren van kennis over gedrag, geeft directe



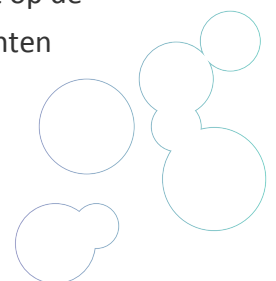
ondervraging over de impact van een interventie veelal een onvolledig beeld (Wilson, 2011). Enerzijds omdat van de deelnemers niet kan worden verwacht dat zij kunnen inschatten hoe zij eraan toe waren geweest indien zij *niet* hadden deelgenomen aan de interventie, anderzijds omdat respondenten vaak geneigd zijn sociaal wenselijke antwoorden te geven. Bij een interview zijn er altijd twee interpretatiefilters: dat van de respondent, die de informatie waarover hij wordt ondervraagd mentaal ordent en communiceert, en dat van de onderzoeker die de verstrekte informatie duidt en analyseert.

Er zijn echter zeker methoden en technieken om bias en misleiding van kwalitatief evaluatieonderzoek tegen te gaan. Het is extra sterk wanneer het onderzoek niet enkel een momentopname is, maar er sprake is van langdurige betrokkenheid (8) en verschillende evaluatiemomenten (*'repeated measures'*) (9). Dat wil zeggen, een periode die lang genoeg is om een adequate representatie te krijgen van de geëvalueerde interventie. Dit stelt de onderzoekers in staat om langere tijd binnen een bepaalde (sub)cultuur te verblijven, vertrouwen te wekken bij respondenten en te controleren voor misinformatie die mogelijk is ontstaan door fouten van de onderzoeker of misverstanden tussen onderzoekers en respondenten. Meerdere onderzoeksmomenten verschaffen bovendien een longitudinale dimensie aan de evaluatie waarmee veranderingen kunnen worden beschreven.

Bovendien kunnen controlesituaties worden gecreëerd, zodat vergelijkingen kunnen worden gemaakt en werkzame factoren van interventies kunnen worden geïdentificeerd (*'counterfactual condition'*). Dit werkt het beste via zogenoemde multi-site studies, waarbij resultaten van verschillende onderzoekslocaties en settings waar de interventie wordt uitgevoerd, worden vergeleken met soortgelijke locaties en settings waar de interventie niet wordt uitgevoerd (10).

Het is essentieel dat een evaluatie *fair dealing* toepast (11). Fair dealing zorgt ervoor dat zo veel mogelijk perspectieven worden toegelaten in het onderzoek. Dus niet enkel dat van de interventiedoelgroep bijvoorbeeld, maar ook van andere betrokkenen zoals uitvoerend professionals. Dit zorgt ervoor dat het gezichtspunt van één enkele groep niet kan worden gepresenteerd als de objectieve representatie van de werkzaamheid van de interventie. Dit is met name van belang voor onderzoek dat gebruik maakt van interviews en andere gesprekstechnieken.

Het is belangrijk dat evaluatieonderzoek zich richt op de *facilitators* en *barriers* van de interventie (12): welke kenmerken van de interventie hadden een (on)gunstig effect op de deelnemers in termen van de doelen die de interventie nastreeft? Welke omslagpunten zijn bij de respondenten te identificeren?



Hoe Algemene methoden dataverzameling te beoordelen? Enkele fictieve voorbeelden

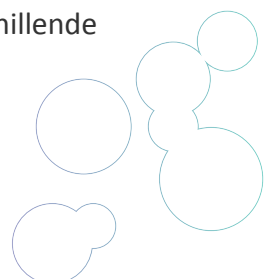
- ✓ In een evaluatie van een straatanpak van overlast gevende jongeren in probleemwijken zijn enkele tientallen interviews afgenomen met leden van jongerengroepen en hangjeugd, maar is niet gesproken met vertegenwoordigers van andere relevante partijen in de wijken zoals de politie of het jeugdwerk. De evaluatie scoort dan 'onvoldoende' op het criterium 'Het onderzoek belicht perspectieven van verschillende groepen die relevant zijn voor de evaluatie ('fair dealing')' (11).
- ✓ In een evaluatie van een voorlichtingscursus om eetstoornissen te voorkomen bij tienermeiden (anorexia-preventie) is via interviews met respondenten aan het licht gebracht dat veel meiden die vatbaar zijn voor anorexia, de tips van de cursus negeren omdat het naleven hiervan een erkenning vereist van hun eigen eetstoornis. De evaluatie identificeerde zodoende gevoelens van zelfstigma en schaamte als barrières voor een effectieve interventie. De evaluatie scoort daarmee 'goed' op het criterium 'Onderzoeksmethoden richten zich op de *facilitators* en *barriers* bij respondenten' (12).

Observatie

13. De evaluatie maakt gebruik van observaties.
14. Observaties zijn gebruikt om voorwaarden van effectiviteit te toetsen ('necessary conditions').
15. Observaties zijn gebruikt om veronderstelde effecten van de interventie te toetsen.
16. Observaties zijn gebruikt om licht te werpen op causale mechanismen van de interventie (oorzaak- en gevolgrelaties werkzame elementen).
17. Interviews en andere gespreksmethoden hadden zoveel mogelijk plaats na gedane observaties (verzamelen van accounts op basis van actions).

Observatiemethoden hebben het voordeel dat verschijnselen en menselijk gedrag uit de eerste hand worden waargenomen. De kracht van observatie is dat op een directe manier een beeld kan worden verkregen van hoe een interventie, programma of methodiek feitelijk gestalte krijgt. Bij interviews kan hiervan achteraf een beeld worden geschetst dat afwijkt van de werkelijke situatie, maar dit geldt niet voor observatie. Een evaluatie die observatie heeft gebruikt (13), wint dus aan validiteit.

Voorts kunnen via observatie voorwaarden voor effectiviteit worden getoetst (14). Veel interventies vereisen bijvoorbeeld een zekere mate van samenwerking tussen verschillende uitvoeringsinstellingen of tussen verschillende belanghebbenden. Niet zelden blijkt

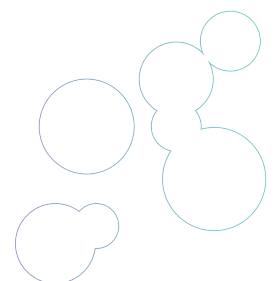


echter achteraf dat de interventie geen effect sorteerde omdat de vereiste samenwerking niet van de grond kwam. Observatiemethoden kunnen licht werpen op de processen die moeten leiden tot de veronderstelde effecten (15) en het kan causale mechanismen (16) achterhalen die hierin van belang zijn.

Interviewmethoden kunnen in samenhang met observaties worden ingezet (17), volgens een werkwijze van *actions* en *accounts* (Marsh, Rosser & Harre, 1978). Door éérst te observeren en daarna te interviewen en niet andersom, voorkomt de evaluator enerzijds dat hij zijn eigen interpretatiekaders te zeer laat domineren in de vraagstellingen en anderzijds dat respondenten hun eigen werkelijkheid construeren over een interventie die wellicht niet strookt met hun eigenlijke gedrag of het effect ervan. De observaties zijn de feitelijke gebeurtenissen die vervolgens als houvast dienen voor het interview. Op basis van gedetailleerde waarneming van de te onderzoeken verschijnselen (onderzoek naar de 'actions') en de verklaringen van de deelnemers voor hun handelen (de 'accounts') wordt een zo compleet mogelijk beeld verkregen.

Hoe Observatie te beoordelen? Enkele fictieve voorbeelden

- ✓ In een evaluatie van een nieuwe gemeentelijke activeringsmethode voor langdurig werklozen zijn 30 interviews gedaan met respondenten van de interventie, maar is slechts één middag geobserveerd bij de uitvoering van de interventie door arbeidsconsulenten. De evaluatie scoort dan 'twijfel' op het criterium 'Het onderzoek maakt gebruik van observaties' (13). Door eenmaal te observeren worden de methodologische voordelen maar minimaal benut. Observatie wint aan kracht bij langdurige betrokkenheid en meerdere onderzoeksmomenten.
- ✓ In een evaluatie van een sportprogramma voor moeilijk opvoedbare jongeren zijn observatiemethoden gebruikt om enkele voorwaarden en veronderstelde effecten van de interventie op waarde te schatten. Zo is nagegaan in hoeverre jongeren zich aanmelden voor het programma, of de juiste doelgroep wordt bereikt en of er aanwijzingen zijn dat jongeren als gevolg van de sportactiviteiten bijvoorbeeld in hun gezin socialer gedrag gaan vertonen (een belangrijk doel van de interventie). De evaluatie scoort dan 'goed' op de criteria 'Observaties zijn gebruikt om voorwaarden van effectiviteit te toetsen ('necessary conditions') (14)' en 'Observaties zijn gebruikt om veronderstelde effecten van de interventie te toetsen' (15).

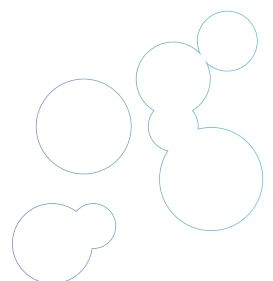


- ✓ In een evaluatie van een training voor werknemers die tot doel heeft collega's met een verstandelijke beperking beter te begeleiden op de werkplek, zijn observaties gebruikt om iets te zeggen over de werkzaamheid van de training. Het evaluatierapport vermeldt echter niet welke schakels in de aanpak leiden tot een betere begeleiding van verstandelijk beperkte werknemers of hierin juist te kort schieten. De evaluatie scoort dan 'onvoldoende' op het criterium 'Observaties zijn gebruikt om licht te werpen op causale mechanismen van de interventie (oorzaak- en gevolgrelaties werkzame elementen)' (16).
- ✓ In een evaluatie van een schuldhulpverleningsaanpak uitgevoerd door sociale wijkteams, zijn 25 interviews afgenomen bij schuldenaren, maar geen observaties gedaan van de uitvoering van de aanpak door sociale professionals van de wijkteams. De evaluatie scoort dan 'onvoldoende' op het criterium 'Interviews hadden zoveel mogelijk plaats na gedane observaties (verzamelen van *accounts* op basis van *actions*)' (17).

Respondentensample

18. Respondenten zijn geselecteerd op basis van relevante kenmerken van de bredere doelgroep(en) van de evaluatie ('purposive sampling').
19. De omvang van het respondentensample en de data die hiermee zijn gegenereerd, zijn afdoende gegeven de interventiekenmerken en het evaluatiedoel.
20. De rapportage bevat een toelichting op eventueel 'gemiste' respondenten (non-participatiedoelgroepen).
21. De onderzoekers hebben gereflecteerd op hoe respondentgedrag of -zienswijzen tijdens de evaluatie door hen mogelijk zijn beïnvloed.

Sampling procedures in kwalitatief onderzoek volgen meestal niet lijnrecht de regels van statistische sampling. Hierbij worden respondenten op basis van algemene kenmerken van een vooraf vastgestelde populatie geselecteerd met het oog op een representatieve steekproef. Dit betekent echter niet dat sampling strategieën in kwalitatief onderzoek niet aan criteria gebonden zijn. Ook bij kwalitatief onderzoek wint de studie aan validiteit als respondenten gericht worden geselecteerd. Alleen gebeurt dit niet zozeer op basis van zuiver statistische representativiteit, maar op basis van relevantie voor het thema dat wordt onderzocht en de situaties die relevant zijn binnen dat onderzoeksthema (Ragin, 1994). De mate waarin de respondenten zijn geselecteerd op basis van relevante kenmerken van de bredere doelgroep(en) van de evaluatie is een criterium ('purposive sampling'). In de praktijk blijkt dit echter niet altijd mogelijk en moeten onderzoekers soms terugvallen op convenience sampling of de sneeuwbal-methode.

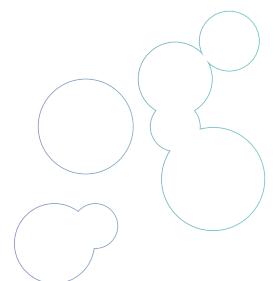


Een evaluatierapport dient niet alleen een beschrijving te geven van de respondenten, maar ook in te gaan op de selectiecriteria die relevant zijn voor de evaluatie en de situaties die daarvoor van toepassing zijn (18). Een evaluatie van een interventie voor gezond eetgedrag bij ouderen bijvoorbeeld, moet wel haar respondenten uit die demografische doelgroep selecteren, en niet bij personen van middelbare leeftijd. Ook de omvang van het respondentensample en de data die hiermee zijn gegenereerd, blijft een relevant criterium. Ondanks het feit dat kwalitatief onderzoek geen statistische representativiteit van een grotere populatie nastreeft, dient het sample afdoende te zijn, gegeven de interventiekenmerken en het evaluatiedoel (19). Een interventie met een groot bereik, dat wil zeggen: een groot aantal deelnemers en betrokkenen, vereist dus ook een wat groter sample (zie ook het tweede voorbeeld in de box hieronder). Als de sample in dit opzicht te klein is, is het bovendien moeilijker een punt van datasaturatie vast te stellen, wat de constructie van conclusies bemoeilijkt (zie onderdeel Analyse en verslaglegging). Een toelichting op eventueel gemiste respondenten is daarbij van belang (20). Non-participatie van respondenten kan namelijk de validiteit van de onderzoeksuitkomsten beïnvloeden door selectiebias en inzichten die onderbelicht blijven.

Reflectie op in beïnvloeding van het respondentgedrag of -zienswijzen tijdens de evaluatie, is eveneens een criterium (21). Anders dan bij veel kwantitatief onderzoek is bij kwalitatief onderzoek directe interactie van de onderzoeker met zijn of haar studieobjecten onvermijdelijk. Het vereist dat de onderzoeker praat met mensen of hen van dichtbij observeert. Dit brengt het risico met zich mee dat respondenten zich laten beïnvloeden of zelfs sturen door de onderzoekers en – bewust of onbewust – hun ervaringen onzuiver weergeven. Dit kan niet altijd worden uitgesloten, maar bij twijfel is reflectie door de onderzoeker noodzakelijk.

Hoe Respondentensample te beoordelen? Enkele fictieve voorbeelden

- ✓ In een evaluatie van een radicaliseringsaanpak zijn tien risicjongeren geïnterviewd. De sampling-wijze was 'convenience' omdat de onderzoekers bij deze doelgroep afhankelijk waren van vrijwillige deelname. De evaluatie scoort dan 'niet van toepassing' op het criterium 'Respondenten zijn geselecteerd op basis van relevante kenmerken van de bredere doelgroep(en) van de evaluatie ('purposive sampling')' Gezien de aard van het thema radicalisering is purposive sampling immers niet mogelijk (18).



- ✓ In een evaluatie van een nationale begeleidings- en resocialisatieaanpak van jeugdige veelplegers zijn voor het onderzoek vijf jongeren geïnterviewd. Bij het programma waren gedurende de looptijd van een jaar landelijk circa 200 delinquente jongeren van uiteenlopende achtergronden betrokken. De evaluatie scoort dan 'onvoldoende' op het criterium 'De omvang van het respondentensample en de data die hiermee zijn gegenereerd, zijn afdoende gegeven de interventiekenmerken en het evaluatiedoel' (19).
- ✓ In een evaluatie van een gemeentelijke aanpak om eenzame ouderen uit hun isolement te halen, zijn enkele tientallen ouderen geïnterviewd. De onderzoekers vermelden echter in hun rapportage dat zij voor de respondentselectie afhankelijk waren van die ouderen die ambtenaren van de gemeente voor hun aanpak wisten te bereiken. Vervolgens bespreken zij welke invloed deze selectie mogelijk heeft gehad op de uitkomsten van de evaluatie. De evaluatie scoort dan 'goed' op het criterium 'De rapportage bevat een toelichting op eventueel 'gemiste' respondenten (non-participatie doelgroepen)' (20).

5.4 Analyse en verslaglegging

Het onderdeel Analyse en verslaglegging toetst de analytische sterkte van het onderzoek en of bepaalde procedures en analysetechnieken zijn toegepast die de validiteit van het onderzoek vergroten. Denk bijvoorbeeld aan de concrete toepassing van validiteitscriteria en -procedures zoals triangulatie, member check, *negative case selection* en *peer debriefing*, maar ook de mate waarin de onderzoekers in hun analyse samenhang hebben aangebracht tussen de onderzoeksvragen, data, analyses en conclusies.

Illustraties resultaten

- 22. Het verslag bevat illustraties van de originele data (bijv. citaten), zodanig dat de lezer een beeld krijgt van de evaluatie en de uitkomsten.
- 23. De onderzoeksresultaten zijn gedetailleerd beschreven aan de hand van citaten, fragmenten en contextbeschrijving ('thick description').

Bij de presentatie van een heldere kwalitatieve analyse is het zaak dat het verslag illustraties bevat van de originele data (bijvoorbeeld citaten), zodanig dat de lezer een beeld krijgt van de evaluatie en de uitkomsten zoals die zich hebben gemanifesteerd in de 'echte' wereld (22). Met het oog hierop dienen de onderzoeksresultaten gedetailleerd beschreven te zijn aan de hand van citaten uit interviews en fragmenten uit observaties (23, 'thick description').

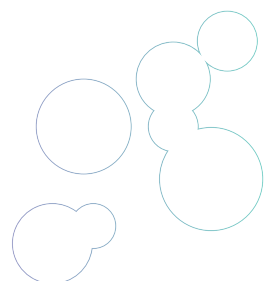


Analysetechnieken en procedures

24. Het verslag bevat een toelichting wanneer een punt van datasaturatie is bereikt en hoe dit is vastgesteld.
25. De analyse/codering van de data is gedaan door verschillende onderzoekers (peer analyse / onafhankelijke codering).
26. Resultaten van het onderzoek zijn gecontrasteerd met andere databronnen ('triangulatie').
27. Interview- of observatieverslagen of resultaten zijn ter controle teruggekoppeld naar respondenten ('member check').
28. Onderzoekers hebben het onderzoeksproces en de keuzes daarin nauwgezet en chronologisch gedocumenteerd ('audit trail').
29. In de data-analyse is bewust naar deviante 'uitschieters' gezocht en deze worden verklaard ('negative case selection').

In tweede instantie kunnen verschillende analysetechnieken en -procedures de validiteit en betrouwbaarheid van kwalitatief evaluatieonderzoek vergroten. De meeste van deze technieken zijn bedoeld om bias van de kant van onderzoekers of onzuivere gevolgtrekkingen te voorkomen. Een evaluatie wint aan validiteit als de onderzoekers toelichten wanneer een punt van datasaturatie is bereikt en hoe dit is vastgesteld (24). Datasaturatie betekent dat de onderzoekers een punt bereiken in de analyse waarbij het samplen van nog meer data niet leidt tot meer of waardevollere informatie in het licht van de onderzoeksvragen. Datasaturatie geeft daarmee een richtpunt wanneer (generaliserende) conclusies over de evaluatieresultaten kunnen worden geconstrueerd en afgerond.

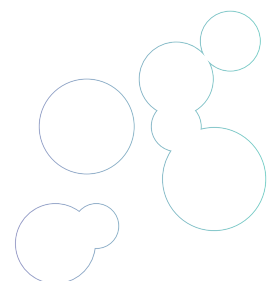
Het komt de analyse ten goede als de codering van de data is gedaan door verschillende onderzoekers (25, peer analyse/onafhankelijke codering) en de resultaten van het onderzoek zijn gecontrasteerd met andere databronnen (26, 'triangulatie'). Triangulatie kan bestaan uit een vergelijking van onderzoekslocaties (bijvoorbeeld uitkomsten van verschillende wijken die met elkaar worden gecontrasteerd), een vergelijking van interviewmateriaal met observaties of een vergelijking van het narratief van respondenten met kwantitatieve onderzoeksresultaten. Een 'member check' houdt in dat feedback wordt verkregen van informanten of respondenten over de verzamelde data en interpretaties en conclusies van de evaluatie (27). Hiermee wordt respondenten en informanten de mogelijkheid gegeven een rol te spelen in het bepalen van de geloofwaardigheid van het onderzoek. In feite gaat het dus om het minimaliseren van de kans op misrepresentatie en misinterpretatie van de onderzochten door de onderzoekers.



Een *'audit trail'* houdt in dat onderzoekers het onderzoeksproces en de keuzes daarin nauwgezet en chronologisch documenteren, bijvoorbeeld via een logboek en memo's (28). Dit documentatiespoor stelt externe beoordelaars – maar ook de onderzoekers zelf – beter in staat vast te stellen of de bevindingen worden gestaafd door de data, of de gevolgtrekkingen logisch zijn en of (methodologische) keuzes in het onderzoek kunnen worden gerechtvaardigd. *'Negative case selection'* is een vorm van data-analyse waarbij de eigen interpretatie van de data zoveel mogelijk op de proef wordt gesteld door bewust naar uitschieters (*'negative cases'*) te zoeken en deze te verklaren (29). Een enkele negative case kan al voldoende zijn om een bepaalde hypothese of verklaring aan te passen.

Hoe Analysetechnieken en procedures te beoordelen? Enkele fictieve voorbeelden

- ✓ In een evaluatie van een dagbestedingsprogramma voor dementerende ouderen binnen ontmoetingscentra zijn door het hele land observaties gedaan, waarbij per locatie verschillende onderzoekers zijn ingezet. De data (observatieverslagen) die dit heeft opgeleverd zijn tussen de onderzoekers uitgewisseld zodat verschillende personen onafhankelijk van elkaar het materiaal hebben gecodeerd. De evaluatie scoort daarmee 'goed' op het criterium 'De analyse/codering van de data is gedaan door verschillende onderzoekers (peer analyse/onafhankelijke codering)' (25).
- ✓ In een evaluatie van een interventie voor vrouwenopvang (huiselijk geweld) is voor het onderzoek gebruik gemaakt van observaties, interviews en documentanalyse (analyse van e-mailcorrespondentie tussen de vrouwen en begeleiders). De resultaten van de verschillende databronnen worden echter grotendeels apart van elkaar besproken en in de conclusies wordt geen poging gedaan de verschillende bronnen analytisch met elkaar te verknopen. De evaluatie scoort dan twijfel op het criterium 'Resultaten van het onderzoek zijn gecontrasteerd met andere databronnen ('triangulatie')' (26).
- ✓ In een evaluatie van een gemeentelijke aanpak om bewonersparticipatie te bevorderen in een herstructureringswijk, zijn enkele tientallen bewoners geïnterviewd over hun ervaringen met de gemeente. Gedurende het onderzoek hebben onderzoekers de weergave van hun interviews ter controle teruggekoppeld aan bewoners. Daarnaast zijn in de laatste fase van het onderzoek meerdere bewoners betrokken bij de constructie van conclusies. De evaluatie scoort daarmee 'voldoende' op het criterium 'Interview- of observatieverslagen of resultaten zijn ter controle teruggekoppeld naar respondenten ('member check')' (27).



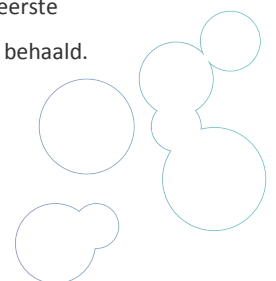
Constructie conclusies en verklaring modellen

30. Op basis van de bevindingen is een verklarend model geconstrueerd voor de (veronderstelde) causale relaties van de interventie.
31. Op basis van de bevindingen wordt een analyse gepresenteerd die iets zegt over de effectiviteit van de interventie.
32. De evaluatieresultaten worden besproken in het licht van de onderzoeksvragen (samenhang resultaten en analyse).
33. De conclusies van de evaluatie zijn coherent en geloofwaardig ('credibility').
34. Conclusies van de evaluatie zijn voorgelegd aan externe deskundigen ('peer debriefing').
35. De onderzoekers gaan in op hun eigen meningen en belangen die van invloed kunnen zijn op de interpretatie van de resultaten (reflexiviteit).

Voor de conclusies en verklaring modellen die een kwalitatieve evaluatie oplevert, is het allereerst zaak – zoals het drempelcriterium aangeeft – dat de resultaten en conclusies van de evaluatie worden besproken in het licht van de interventiedoelen. Het onderzoek moet dus een koppeling maken met de *outcome* criteria die de interventie heeft gesteld. Dit lijkt vanzelfsprekend, maar door persoonlijke voorkeuren, ongewenste resultaten (de 'onwelkome boodschap') of druk van financiers en opdrachtgevers, komt het voor dat onderzoekers evaluatieresultaten weglaten, wegmasseren of in een ander licht bespreken dan de oorspronkelijk geformuleerde vraagstelling⁴.

De evaluatie wint aan validiteit als hierbij op basis van de bevindingen een verklarend model is geconstrueerd voor de (veronderstelde) causale relaties van de interventie (30). Een dergelijk model kan verschillende vormen aannemen. Idealiter worden de causale relaties die de onderzoekers hebben blootgelegd gevisualiseerd (bijvoorbeeld met pijldiagrammen), maar dit is niet per se noodzakelijk. Het kan ook in lopende tekst, als de oorzaak- en gevolgrelaties tussen belangrijke variabelen van de interventie maar duidelijk worden. Op basis van de bevindingen dient dus op zijn minst een analyse te worden gepresenteerd die iets zegt over de effectiviteit van de interventie (31) en er moet tussen deze eindanalyse en de resultaten van de evaluatie een zekere samenhang zijn (32). De conclusies moeten dus, anders gezegd, logisch voortvloeien uit de onderzoeksresultaten en geloofwaardig zijn (33). Geloofwaardigheid (*credibility*) heeft in de context van kwalitatieve analysecriteria vooral betrekking op een rapportage van onderzoek die de observaties en ervaringen van

⁴ Uiteraard kan een evaluatie tevens inzicht geven in de werking van de interventie rond aspecten die niet van tevoren kunnen worden voorspeld. Kwalitatief onderzoek levert vaak ook onvermoede resultaten op over een interventie. De oorspronkelijke doelen worden niet behaald, maar andere doelen misschien wel. Dit laat echter onverlet dat in eerste instantie de evaluatie moet bespreken in hoeverre de *oorspronkelijke* geformuleerde interventiedoelen worden behaald.

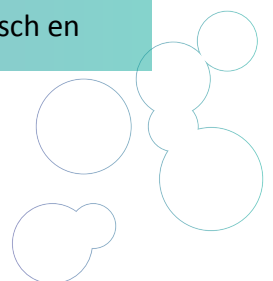


respondenten accuraat representeert. De conclusies moeten dus geloofwaardig zijn gegeven de ruwe resultaten van het onderzoek. Vandaar ook het belang van illustraties van originele data in het verslag.

Ten slotte gelden er nog enkele andere criteria om de betrouwbaarheid en validiteit van de constructie van conclusies te vergroten. Wanneer conclusies van de evaluatie zijn voorgelegd aan externe deskundigen, verkleint dit de kans op bias of weglating (34, 'peer debriefing'). Peer debriefing is een vorm van externe evaluatie van het kwalitatieve onderzoeksproces. Lincoln en Guba (1985: 308) omschrijven de rol van de peer debriefer als 'de advocaat van de duivel'. Het zijn personen die – eventueel in een klankbordgroep – lastige vragen stellen over de betekenissen, interpretaties en conclusies van het onderzoek en die de onderzoekers dwingen tot eerlijkheid. Als de onderzoekers hierbij bovendien ingaan op hun eigen meningen en belangen die van invloed kunnen zijn op de interpretatie van de resultaten, dan verkleint dit eveneens de kans op bias (35, reflexiviteit).

Hoe Conclusies en verklaringsmodellen te beoordelen? Enkele fictieve voorbeelden

- ✓ In een evaluatie van een methodiek gericht op empowerment van mishandelde vrouwen zijn aan de hand van een coherent schema de factoren binnen de interventie beschreven die van invloed zijn op de bevordering van zelfvertrouwen en zelfredzaamheid van de vrouwen, waar deze werkzame elementen van afhankelijk zijn en welke handelingen van professionals die de methodiek uitvoeren de kwaliteiten en mogelijkheden van vrouwen verbeteren. De evaluatie scoort daarmee 'goed' op het criterium 'Op basis van de bevindingen is een verklarend model geconstrueerd voor de (veronderstelde) causale relaties van de interventie' (30).
- ✓ In een evaluatierapport van een methode voor arbeidsactivering van bijstandsontvangers worden weliswaar conclusies besproken, maar er wordt op basis van het onderzoek niet ingeschat of de interventie ook doet waarvoor die bedoeld is. Het rapport bespreekt de ervaringen van de respondenten en professionals, maar gaat niet in op de vraag of de aanpak bijdraagt aan een hogere kans op werkhervatting van de bijstandsontvangers. De evaluatie scoort dan 'onvoldoende' op het criterium 'Op basis van de bevindingen wordt een analyse gepresenteerd die iets zegt over de effectiviteit van de interventie' (31).
- ✓ In een evaluatie van een schoolprogramma om tolerantie tussen LHBTI-ers en andere leerlingen te vergroten, is gedurende het onderzoekstraject eenmalig een externe deskundige geraadpleegd. Er is echter geen formele klankbordgroep of begeleidingscommissie geformeerd die de bevindingen van de evaluatie kritisch en



systematisch tegen het licht heeft gehouden. De evaluatie scoort dan 'twijfel' op het criterium 'Conclusies van de evaluatie zijn voorgelegd aan externe deskundigen ('peer debriefing')' (34).

5.5 Bijdrage en inbedding kennis

Het laatste thema, Bijdrage en inbedding kennis, toetst in hoeverre het onderzoek aansluit bij bestaande kennis over het onderwerp en de implicaties voor de toekomst. Anders dan de methodologische en analytische criteria hebben deze criteria een minder directe toepassing op de validiteit en betrouwbaarheid van een evaluatie. Indirect zijn deze criteria echter wel degelijk van belang voor de beoordeling van een evaluatie. Door aan te sluiten bij eerdere kennis en theorie verschaft het onderzoek zich epistemologische diepgang. Hiermee wordt onder andere voorkomen dat eerder ontwikkelde relevante kennis over het onderwerp over het hoofd wordt gezien en kunnen toekomstige, opvolgende evaluaties beter profiteren van de nieuw ontwikkelde kennis.

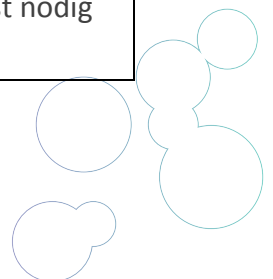
Koppeling literatuur en theorie

36. De vertrekpunten van de evaluatie zijn ingebed in bestaande theorievorming.
37. De resultaten van de evaluatie worden in een bredere context geplaatst door aan te sluiten bij eerder relevant onderzoek/literatuur/theorie.

Vrijwel alle onderwerpen of interventies zijn wel eerder in enigerlei vorm onderzocht. Ook voor nieuw ontworpen interventies geldt vaak dat de vertrekpunten of bepaalde kenmerken van de aanpak al eerder in een andere vorm zijn geïmplementeerd en onderzocht. Het is belangrijk dat de onderzoekers zich hiervan rekenschap geven en hun eigen evaluatie zich hiertoe laten verhouden. Dit betekent ook dat de (methodologische) vertrekpunten van de evaluatie zelf zijn ingebed in bestaande theorievorming (36). Waarom is er gekozen voor een bepaald evaluatieontwerp? Waarom deze specifieke methoden en analyses? Wat maakt dit tot logische keuzes? Tot slot is het essentieel dat na afloop van de evaluatie de onderzoeksresultaten in een bredere context worden geplaatst (37). Ook hiertoe dient te worden aangesloten bij eerder relevant onderzoek of theorievorming in de literatuur.

Kennisontwikkeling en implicaties van de evaluatie

38. De onderzoekers bespreken expliciet welke nieuwe inzichten de evaluatie heeft opgeleverd.
39. De onderzoekers bespreken aan de hand van de resultaten wat er in de toekomst nodig is aan kennisontwikkeling over de interventie (lacunes).



40. De onderzoekers bespreken of en hoe de resultaten kunnen worden gegeneraliseerd naar andere interventiecontexten.

41. De onderzoekers bespreken aan de hand van de resultaten aanbevelingen voor beleid- en praktijk.

De ontwikkelde kennis van de evaluatie is ook relevant voor de toekomst. De onderzoekers dienen te bespreken welke nieuwe inzichten de evaluatie heeft opgeleverd (38) en wat er op basis van de resultaten in de toekomst nodig is aan extra kennisontwikkeling over de interventie (39). Een goede evaluatie beantwoordt namelijk niet alleen vragen, ze formuleert ook nieuwe vragen. Dit betekent bijvoorbeeld dat de onderzoekers bij dit criterium aangeven wat onbekend en onderbelicht blijft over de effectiviteit van de interventie (kennislacunes) en welk type onderzoek er in de toekomst nodig is om deze lacunes op te vullen. Ook is het zinvol dat de onderzoekers bespreken of en hoe de resultaten kunnen worden gegeneraliseerd naar andere interventiecontexten (40). Causale mechanismen of werkzame elementen van specifieke interventies zijn vaak overdraagbaar naar andere interventiecontexten. Een evaluatie in het ene veld kan dus waardevol zijn voor een interventie in een ander veld. Op de lange termijn kan dit bijdragen aan het opbouwen van een meer algemeen kennisreservoir en theoretische basis over ‘families of interventions’ (Pawson, 2006).

Tot slot speelt – zoals eerder besproken bij het onderdeel Context en achtergrond – geen enkele sociale of gedragsinterventie zich af in een maatschappelijk vacuüm.

Evaluatieonderzoek is een vorm van praktijkgericht onderzoek en heeft alles te maken met beleid (Swanborn, 2004). Het is dus zaak dat de onderzoekers aan de hand van hun gevonden resultaten beleids- en praktijkimplicaties bespreken (41).

Hoe Bijdrage en inbedding kennis te beoordelen? Enkele fictieve voorbeelden

- ✓ In een evaluatie van een aanpak in de hulpverlening aan thuisloze jongeren is op basis van een *grounded theory* analyse een theoretisch model geconstrueerd over hoe de aanpak bijdraagt aan de oplossing van dak- en thuisloosheid van jongeren. De evaluatie scoort hiermee ‘goed’ op het criterium ‘De vertrekpunten van de evaluatie zijn ingebed in bestaande theorievorming’ (36).
- ✓ In een evaluatie van een cursus om ouderen meer te laten bewegen, worden de uitkomsten van het onderzoek en of de interventie effectief kan worden bevonden zorgvuldig besproken. De auteurs van het rapport laten echter na te bespreken hoe de uitkomsten kunnen worden verbreed naar (andere) praktische toepassingen van interventies ter bewegingsbevordering van ouderen en wat dit betekent voor toekomstig beleid. De evaluatie scoort daarmee ‘onvoldoende’ op het criterium ‘De onderzoekers bespreken aan de hand van de resultaten aanbevelingen voor beleid- en praktijk’ (41).



6. Toelichting checklist mixed methods onderzoek

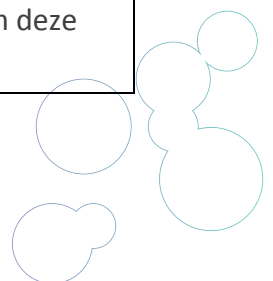
In dit document staat kwalitatief onderzoek centraal. Met betrekking tot effectiviteitsonderzoek naar interventies is het echter niet ongebruikelijk meer dat kwalitatieve en kwantitatieve onderzoeken worden gecombineerd. Bijvoorbeeld open interviews met survey-resultaten of RCT's met observaties. Steeds meer van dergelijk gecombineerd onderzoek vindt zijn weg naar de erkenningstrajecten en databanken van de betrokken instituten. Daarom zijn naast de criteria voor kwalitatief evaluatieonderzoek enkele globale richtlijnen vastgesteld voor *mixed methods* onderzoek. De richtlijnen zijn vevat in een aparte checklist.

Mixed methods onderzoek maakt gebruik van kwalitatief en kwantitatief onderzoek waarvoor op zichzelf aparte methodologische en analytische validiteitscriteria gelden. Wil een mixed methods evaluatie niettemin meerwaarde halen uit het feit dat het een combinatie van methoden hanteert, dan gelden er wel enkele richtlijnen. De checklist doet hiervoor op hoofdlijnen een aanzet. De 11 richtlijnen zijn gebaseerd op inzichten uit de methodologische literatuur (zie vooral Teddlie & Tashakkori, 2009; Lub, 2014; Pope, Mays & Popay, 2007).

De gemeenschappelijke factor van de verschillende richtlijnen in de checklist is het bereiken van een synergie-effect. Met andere woorden, dat *meerwaarde* wordt gecreëerd door het feit dat mixed methods zijn ingezet om de impact van een interventie te toetsen, in plaats van dat enkel voor een bepaalde methode is gekozen. Invulling van de checklist geeft een erkenningscommissie daarmee een indicatie van het niveau van toegevoegde waarde van het ter beoordeling staande mixed methods onderzoek. Anders dan de checklist voor kwalitatief evaluatieonderzoek, moet deze checklist niet worden gelezen als een systematiek met harde criteria in combinatie met een puntentelling.

6.1 Vertrekpunten onderzoek

1. De onderzoekers onderbouwen de keuze voor mixed methods.
2. De kwantitatieve en kwalitatieve onderzoeksmethoden zijn ingezet om gerelateerde aspecten van dezelfde onderzoeksvraag te beantwoorden.
3. Het onderzoek streeft naar een integratie van kwantitatieve en kwalitatieve onderzoeksresultaten en analysetechnieken.
4. Het onderzoek bevat een beschrijving van de analyse-strategie die is gebruikt om deze integratie te bewerkstelligen.



Allereerst is het zaak dat de onderzoekers hun keuze voor mixed methods onderbouwen (1). Waarom is er gekozen voor een meervoudige aanpak? Waarom juist deze combinatie van methoden, en niet een andere? Welke informatie of meerwaarde verwachtte men hieruit te halen? Ten tweede is het belangrijk dat de kwantitatieve en kwalitatieve onderzoeksmethoden zijn ingezet om gerelateerde aspecten van *dezelfde* onderzoeksvraag te beantwoorden (2). De premisse van mixed methods onderzoek is dat het betere kennis oplevert over het onderzochte verschijnsel dan een keuze voor enkel kwantitatieve dan wel kwalitatieve methoden. Maar om dit kennisniveau te bereiken, moeten de uitkomsten van de twee onderzoeklijnen (kwantitatief/kwalitatief) wel effectief met elkaar worden verbonden en niet los van elkaar staan. Het onderzoek moet hierbij streven naar een integratie van kwalitatieve en kwantitatieve onderzoeksresultaten en analysetechnieken (3). Concreet houdt dit in dat in het evaluatierapport de onderzoeksvragen en -resultaten min of meer gelijktijdig en in onderlinge samenhang worden beantwoord en beschreven. Tot slot is het van belang dat het onderzoek een beschrijving bevat van de analysestrategie die is gebruikt om deze integratie te bewerkstelligen (4). Mixed methods evaluatieonderzoek kan op verschillende designs gefundeerd zijn (zie Teddlie & Tashakkori, 2009) en elk design legt verschillende accenten en vereist wat andere analysetechnieken. Dit analytisch kader moet duidelijk omschreven zijn.

6.2 Resultaatanalyse

5. De resultaten van de kwalitatieve en kwantitatieve onderzoeklijnen zijn in samenhang geanalyseerd, gecontrasteerd en gerapporteerd ('triangulatie').
6. Resultaten van beide onderzoeklijnen krijgen in de rapportage evenveel prioriteit.
7. Het onderzoek expliciteert wanneer en hoe bepaalde uitkomsten van de verschillende onderzoeklijnen zijn gebruikt en hoe zij zich tot elkaar verhouden.
8. Het onderzoek expliciteert welke onderzoekers verantwoordelijk waren voor welke methoden en analyses en hoe zij met elkaar afstemden.

Kijken we naar de richtlijnen voor resultaatanalyse van mixed methods onderzoek, dan is het allereerst zaak dat – synchroon met de onderzoeksvragen – de resultaten van de kwalitatieve en kwantitatieve onderzoeklijnen in samenhang zijn geanalyseerd, gecontrasteerd en gerapporteerd (5). Het moge duidelijk zijn dat mixed methods onderzoek een integratie vereist van zowel statistische als tekstuele analysetechnieken en meer bijzondere toepassingen zoals triangulatie en 'dataconversie' (het transformeren van kwalitatieve gegevens naar kwantitatieve gegevens of vice versa). Het zorgvuldig combineren en vergelijken van verschillende databronnen, dataverzamelmethode en onderzoeksprocedures is dus een belangrijk vereiste. Cruciaal hierbij is dat de resultaten van beide onderzoeklijnen evenveel prioriteit krijgen in de rapportage (6). Het is dus niet de bedoeling dat de ene methode 'ondergeschikt' wordt gemaakt aan de andere.



Geëxpliciteerd dient te worden wanneer en hoe bepaalde uitkomsten van de verschillende onderzoekslijnen zijn gebruikt en hoe zij zich tot elkaar verhouden (7). Niet zelden worden bij evaluaties die gebruik maken van een combinatie van methoden de kwalitatieve uitkomsten in de eindrapportage slechts zijdelings besproken – of zelfs helemaal niet – en gaat alle aandacht uit naar de cijfers, bijvoorbeeld percentages van de surveyresultaten, significantieniveaus of de effectgrootte van een RCT.

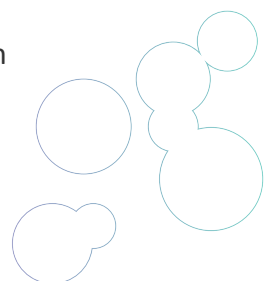
Tot slot is van het belang dat het onderzoek expliciteert welke onderzoekers verantwoordelijk waren voor welke methoden en analyses en hoe zij in dit proces met elkaar afstemden (8). De analysetechnieken voor beide onderzoekslijnen vereisen vaak aparte specialismen en specifieke methodologische kennis van de onderzoekers. Dus is het ook – met het oog op de kwaliteit van het onderzoek – belangrijk dat de analyse van het kwantitatieve en kwalitatieve deel van het onderzoek die specialismen en verschillen in methodologische vakbekwaamheid in acht neemt en expliciteert.

6.3 Conclusies

9. Het onderzoek levert verklarende informatie op over de geëvalueerde interventie.
10. Conclusies die voortkomen uit beide onderzoekslijnen worden aangewend om metaconclusies te trekken.
11. De onderzoekers bespreken welke nieuwe inzichten de mixed methods benadering heeft opgeleverd.

Voor het conclusiedeel van een mixed methods evaluatie dient het onderzoek verklarende informatie op te leveren over de geëvalueerde interventie (9). Het voordeel van een mixed method design is tenslotte dat zij een rijker geschakeerd beeld oplevert over de werking en impact van een interventie dan alleen een kwantitatieve of een kwalitatieve methode. Dit heeft ook implicaties voor de constructie van conclusies. In het gunstigste geval komen uit het onderzoek conclusies voort van een ‘hogere conceptuele orde’ over de geëvalueerde interventie (Teddlie & Tashakkori, 2009: 286). Het rapport bevat dus idealiter niet enkel uitspraken over of de interventie werkt, maar ook hoe die werkt, voor welke doelgroepen en onder welke omstandigheden. Conclusies die voortkomen uit beide onderzoekslijnen moeten dus worden aangewend om metaconclusies te trekken (10). In de rapportage moeten de onderzoekers duidelijk beargumenteren hoe zij via een dergelijk meervoudige analyse tot hun conclusies zijn gekomen.

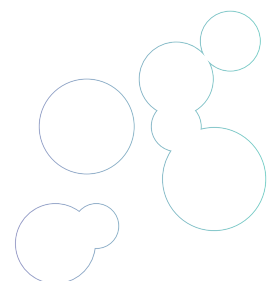
Tot slot is het zaak dat de onderzoekers bespreken welke nieuwe inzichten de mixed methods benadering heeft opgeleverd (11). Dit betekent bijvoorbeeld ook dat de onderzoekers aangeven wat onbekend en onderbelicht blijft over de effectiviteit van



de interventie (kennislacunes) en welk onderzoek er in de toekomst nodig is om deze lacunes op te vullen.

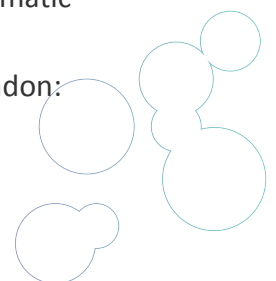
Hoe mixed methods onderzoek te beoordelen? Enkele fictieve voorbeelden

- ✓ In een evaluatie van een gezondheidstraining voor ouderen is een focusgroep georganiseerd met een selectie van deelnemers en een kwantitatieve *survey* uitgezet onder alle deelnemers van de training. De focusgroep lijkt vooral te zijn gebruikt om stellingen te formuleren voor de survey. In het onderzoeksrapport worden enkel de uitkomsten van de kwantitatieve survey besproken. Verdere inzichten uit de focusgroep over de impact van de training worden niet besproken. De evaluatie scoort daarmee 'onvoldoende' op de richtlijn 'De resultaten van de kwalitatieve en kwantitatieve onderzoekslijnen zijn in samenhang geanalyseerd, gecontrasteerd en gerapporteerd' (5).
- ✓ In een evaluatie van een activeringsprogramma voor bijstandsontvangers (doel: werkhervatting) is een design gebruikt met een RCT en observaties. Voor de RCT is een deel van de bijstandsontvangers willekeurig geplaatst op een wachtlijst, een ander deel is willekeurig geselecteerd voor deelname aan het programma. Ter observatie van het programma hebben onderzoekers contactmomenten tussen arbeidsconsulenten en cliënten bijgewoond. In het onderzoeksrapport gaat alle aandacht uit naar de effectgrootte en significantieniveaus verkregen uit de RCT, maar wordt nauwelijks aandacht besteed aan inzichten uit de observaties. Er worden geen fragmenten of citaten van deelnemers gepresenteerd en inzichten uit de observaties worden niet gebruikt ter verdieping van de cijfermatige uitkomsten van het programma. De evaluatie scoort daarmee 'onvoldoende' op de richtlijn 'Resultaten van beide onderzoekslijnen krijgen in de rapportage evenveel prioriteit' (6).
- ✓ In een evaluatie van een mentoringprogramma voor kansarme jongeren zijn interviews afgenomen met een selectie van jongeren (*mentees*) en is een survey uitgezet onder alle mentoren en mentees van het programma. De uitkomsten van de interviews en de survey zijn in de rapportage zorgvuldig naast elkaar gezet en met elkaar vergeleken. Via een codering van het tekstuele materiaal zijn bovendien bepaalde cijfermatige uitkomsten van de survey van betekenis voorzien. Via deze weg is informatie naar boven gekomen over wanneer het programma de maatschappelijke kansen voor jongeren vergroot, voor welke jongeren dit het meest geldt en onder welke omstandigheden. De evaluatie scoort daarmee 'goed' op de richtlijn 'Het onderzoek levert verklarende informatie op over de geëvalueerde interventie' (9).



7. Bibliografie

- Attree, P. & Milton, B. (2006). Critically appraising qualitative research for systematic reviews: defusing the methodological cluster bombs. *Evidence & Policy*, 109-126.
- Creswell, J. & Miller, D. (2000). Determining Validity in Qualitative Inquiry. *Theory into Practice*, 124-130.
- Latour, B. (1988). *Wetenschap in actie. Wetenschappers en technici in de maatschappij*. Amsterdam: Bert Bakker.
- Lincoln, Y. & Guba, E. (1985). *Naturalistic Inquiry*. Thousand Oaks: Sage.
- Lub, V. (2014). *Kwalitatief evalueren in het sociale domein. Mogelijkheden en beperkingen*. Den Haag: Boom Lemma uitgevers.
- Lub, V. (2020). *Kwalitatief evaluatieonderzoek op waarde schatten. Een checklist met passende criteria*.
- Marsh, P., Rosser, E. & Harre, R. (1978). *The Rules of Disorder*. London: Routledge.
- Mays, N. & Pope, C. (2000). Qualitative research in health care. Assessing quality in qualitative research. *BMJ*, 50-52.
- Munn, Z., Porritt, K., Lockwood, C., Aromataris, E. & Pearson, A. (2014). Establishing confidence in the output of qualitative research synthesis: the ConQual approach. *BMC Medical Research Methodology*.
- Omlo, J., Bool, M. & Rensen, P. (2013). *Weten wat werkt. Passend evaluatieonderzoek in het sociale domein*. Amsterdam: SWP.
- Pawson, R. (2006). *Evidence-based Policy. A Realist Perspective*. London: Sage Publications.
- Pope, C., Mays, N. & Popay, J. (2007). *Synthesizing Qualitative and Quantitative Health Evidence. A guide to methods*. Berkshire: McGraw Hill.
- Ragin, C. (1994). *Constructing Social Research*. Thousand Oaks: Pine Forge Press.
- Rychetnik, L., Frommer, M., Hawe, P. & Shiell, A. (2002). Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health*, 119-127.
- Spencer, L., Ritchie, J., Lewis, J. & Dillon, L. (2003). *Quality in Qualitative Evaluation: A framework for assessing research evidence*. London: National Centre for Social Research.
- Swanborn, P. (2004). *Evalueren. Het ontwerpen, begeleiden en evalueren van interventies: een methodische basis voor evaluatie-onderzoek*. Amsterdam: Boom Onderwijs.
- Teddlie, C. & Tashakkori, A. (2009). *Foundations of Mixed Methods Research*. Thousand Oaks: Sage.
- Tong, A., Sainsbury, P. & Craig, J. (2007). Consolidated criteria for reporting qualitative research (COREQ): a 32-item checklist for interviews and focus groups. *International Journal for Quality in Health Care*, 349-357.
- Waterman, H., Tillen, D., Dickson, R. & Koning, K. de (2001). Action research: a systematic review and guidance for assessment. *Health Technology Assessment*.
- Wilson, T. (2011). *Redirect. The Surprising New Science of Psychological Change*. London: Penguin Books.



8. Verantwoording

8.1 Aanleiding en opdracht

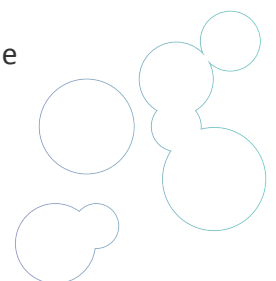
Zes kennisinstituten (Movisie, Nederlands Jeugdinstituut, RIVM Gezond Leven, Trimbos-Instituut, Kenniscentrum Sport en Bewegen, Vilans) hebben zich verenigd in het erkenningstraject. Negen (deel)erkenningcommissies kijken naar de kwaliteit en effectiviteit van interventies op het terrein van bewegen & sport, gezondheidsbevordering, jeugdgezondheidszorg, jeugdzorg/brede jeugdhulp, ontwikkelingsstimulering, langdurende geestelijke gezondheidszorg, ouderenzorg, gehandicaptenzorg, justitie en het sociaal domein. Erkende interventies worden geplaatst in de databanken effectieve interventies. Bij het beoordelen van interventies wordt gekeken naar de resultaten van kwantitatief en kwalitatief onderzoek. Vanuit verschillende commissies kwam de vraag: wat is goed kwalitatief evaluatieonderzoek? En wanneer is het betrouwbaar en valide? Welke criteria kunnen we daarbij aanleggen?

Namens de zes instituten in het erkenningstraject gaf Movisie opdracht aan het Bureau voor Sociale Argumentatie van onderzoeker Vasco Lub om hier onderzoek naar te doen en dit te vertalen in een praktisch toepasbaar instrument voor beoordeling. Het resultaat zijn twee checklists met toelichting: één voor kwalitatief evaluatieonderzoek algemeen en één voor *mixed methods*, een gecombineerd onderzoek met kwalitatieve en kwantitatieve (cijfermatige) elementen. Deze checklists zijn met uitgebreide verantwoording bijeengebracht in de publicatie *Kwalitatief evaluatieonderzoek op waarde schatten. Een checklist met passende criteria* (Lub, 2020).

8.2 Ontwikkeling kwalitatief evaluatieonderzoek

Deze checklists passen in de ontwikkeling die kwalitatief evaluatieonderzoek doormaakt in jeugdhulp, zorg, preventie, sociaal domein en sport. Die ontwikkeling is drieledig. Ten eerste is er het groeiende besef dat kwalitatief onderzoek een zelfstandige rol heeft te spelen bij het verkrijgen van kennis over wat werkt. Denk bijvoorbeeld aan inzichten in veronderstellingen of geïmpliceerde causale verbanden binnen interventies.

Ten tweede ontbreekt een standaard, een ijkpunt voor niet-kwantitatieve onderzoeksvormen. Voor kwantitatieve en experimentele data verkregen uit bijvoorbeeld *surveys* of randomized controlled trials (RCT's) is zo'n ijkpunt wel beschikbaar. Voor kwalitatieve methoden ontbreekt het aan een toetsingsformat. Dit kan leiden tot een zekere willekeur binnen de beoordeling. Des te belangrijker is het om stil te staan bij de betrouwbaarheid en validiteit van kwalitatief evaluatieonderzoek. Conclusies over de



effectiviteit van interventies of over werkzame elementen van aanpakken hebben immers weinig waarde als de onderzoeksresultaten waarop ze gebaseerd zijn niet valide of betrouwbaar zijn. Evenzeer geldt dit in omgekeerde richting: hoe geldiger de uitkomsten van bepaald kwalitatief evaluatieonderzoek, hoe groter de bewijskracht voor de werkzaamheid van de onderzochte interventie.

Ten derde zijn veel interventies om diverse redenen minder geschikt om kwantitatief dan wel experimenteel te evalueren, de kwestie van 'passend onderzoek' (Omlo, Bool & Rensen, 2013). Niet elk interventie-effect kan in een numerieke uitkomstmaat worden verrat. Wanneer kwalitatief onderzoek niet wordt meegenomen bij de beoordeling van wat werkt, kunnen deze interventies nooit boven het erkenningsniveau 'goed onderbouwd' of 'goed beschreven' uitstijgen.

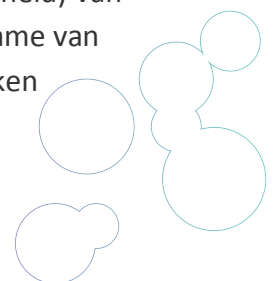
8.3 Constructie checklists

Voor de constructie van de checklists zijn verschillende werkzaamheden verricht. Ten eerste is conceptuele kennis over de betrouwbaarheid en validiteit van kwalitatief onderzoek doorontwikkeld en vertaald naar instrumentele kennis, via een (her)analyse van inzichten uit *Kwalitatief evalueren in het sociale domein* (2014) en de methodologische literatuur die hierin is aangehaald. Ten tweede zijn bestaande algemene (internationale) checklists van kwalitatief onderzoek verzameld en geanalyseerd. Ten derde zijn inzichten over kwalitatieve evaluatie en toetsingsstandaarden bij internationale gezaghebbende instituten verzameld en geïnventariseerd. Dit leverde een aanvullende analyse op, gebaseerd op de door de instituten geleverde input en documenten. Tot slot zijn ter praktische validering van de checklists extra werkzaamheden verricht.

Doorontwikkeling van conceptuele kennis

Kwalitatief evalueren in het sociale domein (Lub, 2014) doet een aanzet voor criteria voor de beoordeling van kwalitatief evaluatieonderzoek. Uit het boek zijn twee sporen te destilleren die kunnen bijdragen aan een zuivere beoordeling van kwalitatief onderzoek. Het eerste is een methodologisch spoor. Bepaalde kwalitatieve methoden en toepassingen van die methoden geven een betrouwbaarder beeld van de werkzaamheid van een interventie dan andere methodologische toepassingen. Dit heeft bijvoorbeeld betrekking op: het toetsen van veronderstellingen over de werking en voorwaarden voor effectiviteit via observatie, een onderzoeksmethodologie volgens *actions* en *accounts* die narratieve misleiding en bias inperkt, en mogelijkheden voor 'quasi-experimenteel' observatieonderzoek waarbij controlesituaties worden gecreëerd in het veld die meer licht werpen op causale relaties.

Het tweede spoor betreft een analytische benadering. Kritiek op de validiteit (geldigheid) van kwalitatief onderzoek heeft sinds de jaren tachtig geresulteerd in een enorme toename van validiteitscriteria en -procedures. Toepassing van die procedures en analysetechnieken kunnen de geldigheid van kwalitatief evaluatieonderzoek aanzienlijk vergroten.



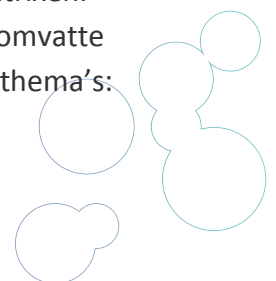
Onderzoekers kunnen voor de validiteitsbeoordeling hun eigen perspectief gebruiken (bijvoorbeeld via triangulatie), dat van participanten in het onderzoek (bijvoorbeeld via *member checks*) of van externe reviewers (bijvoorbeeld via *peer debriefing*).

Voor de constructie van de checklist kwalitatief evaluatieonderzoek is nagegaan welke methodologische en analytische concepten de validiteit van kwalitatief evaluatieonderzoek verhogen en hoe een en ander kan worden vertaald naar criteria, zodanig dat het een praktische manier van beoordelen door externe deskundigen mogelijk maakt. Als bronnen dienden hiertoe inzichten uit *Kwalitatief evalueren in het sociale domein* (Lub, 2014) en een (her)analyse van de methodologische literatuur die hierin is aangehaald. De focus lag op een omzetting van concepten naar criteria binnen een schematisch beoordelingsstelsel dat alternatieven en keuzes in het beoordelingsproces over de betrouwbaarheid en validiteit van het onderzoek expliciteert. Concreet is als beoordelingsvorm gekozen voor de constructie van een checklist waarmee een externe reviewer toetst in hoeverre methodologische toepassingen en analytische procedures zijn gevolgd, resulterend in een score van de validiteit/betrouwbaarheid van het desbetreffende onderzoek.

Analyse van bestaande (internationale) checklists

Als aanvulling op de conceptuele zoektocht en vertaling van methodologische en analytische concepten, zijn bestaande (nationale of internationale) checklists van kwalitatief onderzoek verzameld. Dit leverde 18 afzonderlijke bronnen op (zie bijlage 1). De checklists zijn afkomstig van instituten in zorg en welzijn, gezondheidsbevordering en de geneeskunde, maar zijn vaak ook afkomstig uit de wetenschappelijke methodologische literatuur. Hoewel de checklists een goede aanzet doen om de robuustheid van kwalitatief onderzoek te beoordelen, bracht de analyse ook tekortkomingen aan het licht. Veel van de huidige checklists hanteren vrij multi-interpretabele criteria met open vraagstellingen (vaak zonder toelichting van begrippen) en gebruiken geen scoretoelichting van de criteria. Bovendien zijn ze vooral gericht op toetsing van externe validiteit in plaats van interne validiteit van het onderzoek. Dat wil zeggen, de checklists zijn veelal gericht op criteria die iets zeggen over de generaliseerbaarheid van de uitkomsten naar contexten buiten het onderzoek en of gebruikte methoden duidelijk zijn in de rapportage. Minder vaak gaan ze in op criteria die een zuivere analyse en gevolgtrekking kunnen verbeteren of bias in interpretatie inperken bij 'wat werkt' inschattingen. Daarnaast blijven veel checklists wat voorzichtig in hun uiteindelijke toetsing. Ze stellen weliswaar vragen over kwalitatieve methodologie of analyse, maar stellen daarin geen standaard of ijkpunt vast van methoden of analysetechnieken die extra nadruk verdienen met het oog op de beoordeling van validiteit en betrouwbaarheid.

Alle verzamelde checklists zijn geanalyseerd door de criteria onder te brengen in matrixen. Alle criteria van alle checklists zijn vervolgens thematisch geanalyseerd in Excel. Dit omvatte in totaal 102 afzonderlijke criteria. De criteria zijn vervolgens gerangschikt naar vier thema's: (1) context en achtergrond onderzoek, (2) methodologie en data, (3) analyse en



verslaglegging en (4) bijdrage en inbedding kennis (tevens de hoofdthema's van de nieuwe checklist kwalitatief evaluatieonderzoek). Deze werkwijze verschafte een overzicht van welke criteria hoe vaak gebruikt worden, hoe vigerend die zijn, waarom en op welke manier ze gebruikt worden en waar de hiaten liggen met het oog op validiteit en betrouwbaarheid van kwalitatief evaluatieonderzoek. De input van de conceptuele heranalyse van de methodologische literatuur fungeerde samen met de analyse van de bestaande (internationale) checklists als basis voor de constructie van de nieuwe checklists.

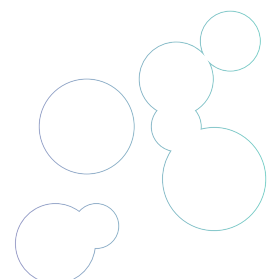
Internationale verkenning en correspondentie

Parallel aan de doorontwikkeling van conceptuele kennis en de verzameling van bestaande algemene checklists, bestond een deel van het project uit een verkenning van internationale trajecten die met het erkenningstraject vergelijkbaar zijn. In verschillende landen en in verschillende sectoren (bijvoorbeeld zorg, gezondheid, welzijn) zijn instituten actief die kwalitatief onderzoek meenemen in hun beoordeling van interventies, of hier in elk geval de kansen toe verkennen. Hun ervaringen en inzichten bleken de moeite waard om mee te nemen in dit traject. Bijvoorbeeld als het gaat om welke criteria essentieel zijn en welke beoordelingsformats werkbaar zijn, maar ook als gaat om het verzamelen van bijvoorbeeld veelbelovende en inspirerende kwalitatieve evaluatieonderzoeken of nieuwe inzichten in het combineren van kwalitatieve methoden met kwantitatieve effectmeting (*mixed methods*).

Gedurende het project is gecorrespondeerd met verschillende internationale instituten en deskundige contactpersonen: zie bijlage 2 voor een overzicht. De contactpersonen van de instituten is vooral gevraagd hoe hun checklist gebruikt wordt, door wie en wat hun ervaringen hiermee zijn. De deskundigen zijn meer inhoudelijk bevroegd op methodologische en analytische kwesties. Tevens zijn aan hen kritiekpunten voorgelegd inzake de tekortkomingen of hiaten in bestaande checklists. Er is over het algemeen positief gereageerd op het project. De correspondentie met deskundigen bevestigde tevens de noodzaak voor de constructie van een aparte checklist specifiek gericht op kwalitatief evaluatieonderzoek, in plaats van op algemeen kwalitatief onderzoek.

Praktische validering van de checklists

Tot slot zijn ter praktische validering van de checklists extra werkzaamheden verricht specifiek gericht op het erkenningstraject. Zo is er steeds afgestemd met de betrokken instituten van het Samenwerkingsverband Erkenning van Interventies over de voortgang van de checklists, zijn verschillende bijeenkomsten bijgewoond van erkenningscommissies van de kennisinstituten en is een aparte klankbordcommissie geformeerd van deskundigen (zie bijlage 3). Conceptversies van de checklists zijn bovendien getest op bruikbaarheid. In de testfase zijn onder andere onderzoeken gebruikt die staan vermeld in de interventie-databanken van de betrokken instituten. Er zijn vier bijeenkomsten van erkenningscommissies bijgewoond: Justitiële Interventies (NJI, Trimbos-instituut en



Movisie), Gezondheidsbevordering (RIVM), Gehandicaptenzorg (Vilans) en Maatschappelijke Ondersteuning, Participatie en Veiligheid (Movisie). Tijdens de bijeenkomsten is het project toegelicht aan de commissieleden, is geobserveerd hoe de commissies interventies beoordelen op basis van onderzoek en is input van de leden over de rol van kwalitatief onderzoek in het erkenningstraject genoteerd. De bijeenkomsten bleken nuttig om bij te wonen. Ze gaven een goed beeld van hoe er wordt beoordeeld, wat daarbij komt kijken en hoe een checklist inzake kwalitatief evaluatieonderzoek de erkenningscommissies kan helpen bij de toetsing van interventies.

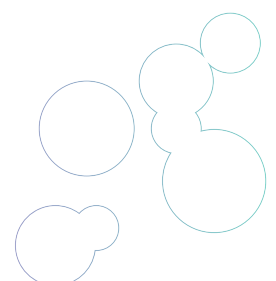
De concept checklists zijn getest in de erkenningscommissies. Dit leverde feedback op over de inhoud en het proces van toetsing. Deze feedback is geïnventariseerd en besproken met projectleiders van het Samenwerkingsverband Erkenning van Interventies en een aantal leden van erkenningscommissies, met als doel de checklists geschikt te maken voor toepassing in het erkenningstraject (zie bijlage 4). Dit heeft tot een aantal wijzigingen geleid die in deze checklists voor het erkenningstraject zijn verwerkt.

Start implementatie

Het Samenwerkingsverband Erkenning van Interventies heeft een implementatieplan gemaakt voor de versterking van de beoordeling van kwalitatief evaluatieonderzoek in het erkenningstraject. Komend jaar worden de checklists stapsgewijs ingevoerd als toetsingsinstrument voor kwalitatief evaluatieonderzoek. Begeleiders van interventieontwikkelaars bij de kennisinstituten, interventie-eigenaren, en natuurlijk commissieleden van de diverse erkenningscommissies, worden desgewenst gefaciliteerd om de tips en tricks van kwalitatief evaluatieonderzoek verder te leren kennen, en toe te passen. We doen dit in een transparant proces waar voor alle partijen leren en verbeteren voorop staat.

De checklists (met toelichting) zijn niet alleen bruikbaar voor beoordeling van interventies in het erkenningstraject. De checklists zijn geschikt voor iedereen die wil bijdragen aan de validiteit en de betrouwbaarheid van kwalitatief evaluatieonderzoek: praktijk, beleid en wetenschap. Het kan ook worden aangewend als een standaard voor kwalitatief evaluatieonderzoek en mixed methods onderzoek. De checklists geven onderzoekers, ontwikkelaars en anderen die te maken hebben met het beoordelen van kwalitatief onderzoek een overzicht van methodologische en analytische criteria die kwalitatief onderzoek naar een hoger niveau kunnen tillen.

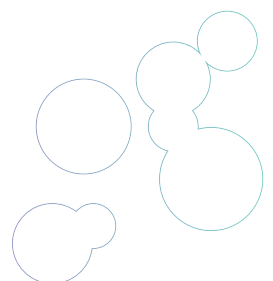
Uiteindelijk staat dit alles ten dienste van de kwaliteitsverbetering van het uitvoerend werk zodat cliënt en burger beter worden ondersteund. Goed uitgevoerd kwalitatief onderzoek kan ons iets leren over welke betekenis (beleids)interventies voor burgers, cliënten, professionals en andere betrokkenen hebben.



Bijlagen

Bijlage 1 Verzamelde (internationale) checklists voor kwalitatief onderzoek

1. Quality assessment of qualitative studies (Bolton, 1996)
2. Cochrane Nederland kwalitatief onderzoek checklist (2013)
3. Schemata for appraising qualitative/interpretive evidence (Rychetnik, Frommer, Hawe & Shiell, 2002)
4. Quality in Qualitative Evaluation: A Framework for Assessing Research Evidence (Spencer, Ritchie, Lewis & Dillon, 2003)
5. Quality Appraisal Checklist (Attree & Milton, 2006)
6. COREQ checklist (Tong, Sainsbury & Craig, 2007)
7. Action research guidance for assessment (Waterman, Tillen, Dickson & De Koning, 2001)
8. Assessing quality in qualitative research (Mays & Pope, 2000)
9. CASP Critical Appraisal (2013)
10. CEBM Critical Appraisal of a Qualitative Study
11. ConQual Approach (Munn, Porritt, Lockwood, Aromataris & Pearson, 2014)
12. GRADE-CERQual (2015)
13. Determining validity in Q research (Creswell & Miller, 2000)
14. Evidence Review Methods, What Works Centre for Wellbeing (2017)
15. ETQS (2002)
16. JBI tool (2017)
17. Mixed Methods Appraisal Tool (2011)
18. SURE Critical Appraisal Qualitative Studies (2018)

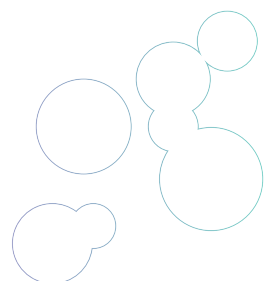


Bijlage 2 Geraadpleegde deskundigen

- Dr. Simon Lewin, Cochrane Norway / Global Health Unit
- Prof. dr. Catherine Pope, University of Southampton
- Dr. Matt Barnard, Natcen U.K.
- Dr. Kritika Samsi, King's College London
- Drs. Djoeke van Dale, senior adviseur, RIVM
- Henk Smid, directeur, ZonMw
- Dr. Karlijn Leenaars, adviseur, RIVM
- Drs. Anne Addink, onderzoeker, NJi
- Dr. Chris Kuiper, lector transformaties in de zorg voor jeugd, Hogeschool Leiden
- Dr. Inge Bastiaanssen, senior adviseur, NJi
- Dr. Marjolein Herps, senior onderzoeker Vilans
- Drs. Robert Gelinck, senior adviseur, Kenniscentrum Sport & Bewegen
- MSc. Evelien Janssen, adviseur, NCJ
- Drs. Thea Meinema, senior adviseur, Movisie
- Drs. Martijn Bool, senior adviseur, Movisie
- Drs. Peter Rensen, programmaleider, Movisie
- Dr. Renske van der Zwet, senior onderzoeker, Movisie
- Drs. Anouk Visser, onderzoeker, Movisie

Bijlage 3 Klankbordcommissie

- Dr. Hennie Boeije, Programmaleider zorg en participatie bij chronische aandoeningen, Nivel
- Dr. Imrat Verhoeven, Assistant Professor of Public Policy, Universiteit van Amsterdam
- Drs. Marco Bik, coördinator kennis en onderzoek, Gemeente Rotterdam BCO-OBI.
- Prof. Dr. Alice Schippers, directeur Disability Studies Nederland en lid van de erkenningscommissie Gehandicaptenzorg
- Drs. Anne Addink, onderzoeker, NJi



Bijlage 4 Geraadpleegde deskundigen mbt. toepassing erkenningstraject

- Prof. Dr. Bram Orobio de Castro, hoogleraar orthopedagogiek en voorzitter van de erkenningscommissie Jeugdzorg en psychosociale/pedagogische preventie
- Drs. Marie-Christine van der Veldt, senior onderzoeker en projectleider NJi in Samenwerkingsverband Erkenning van Interventies en secretaris Erkenningscommissie Justitiële Interventies
- Drs. Djoeke van Dale, senior adviseur en projectleider RIVM Gezond Leven in Samenwerkingsverband Erkenning van Interventies
- Prof. Dr. Ramón Lindauer, hoogleraar Kinder- en Jeugdpsychiatrie en (voorheen) lid van de erkenningscommissie Jeugdzorg en psychosociale/pedagogische preventie
- Dr. Remo Mombarg: lector Bewegingsonderwijs en jeugdsport en docent gedragswetenschappen. Tevens lid erkenningscommissie Sport en bewegen
- Drs. Peter Rensen, programmaleider Effectiviteit en vakmanschap Movisie. Projectleider in Samenwerkingsverband Erkenning van Interventies en tevens (voorheen) secretaris van de Erkenningscommissies Maatschappelijke ondersteuning, participatie en veiligheid en de commissie Justitiële Interventies
- Drs. Nada de Groot, senior onderzoeker Movisie en medewerking aan vooronderzoek checklists
- Drs. Lone von Meyenfeldt, onderzoeker Movisie en secretaris van de erkenningscommissie Maatschappelijke ondersteuning, participatie en veiligheid

