UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Programma-integriteit en effectiviteit van EQUIP voor jongeren in justitiële jeugdinrichtingen in Nederland

Helmond, P.; Brugman, D.; Overbeek, G.

Publication date
2014
Document Version
Final published version
Published in
Pedagogiek

Link to publication

Citation for published version (APA):
Helmond, P., Brugman, D., & Overbeek, G. (2014). Programma-integriteit en effectiviteit van EQUIP voor jongeren in justitiële jeugdinrichtingen in Nederland. *Pedagogiek*, *34*(1), 5-24. http://www.pedagogiek-online.nl/index.php/pedagogiek/article/viewFile/URN%3ANBN%3ANL%3AUI%3A10-1-100774/465

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:15 Sept 2026

Onderzoek

Programma-integriteit en effectiviteit van EQUIP voor jongeren in justitiële jeugdinrichtingen in Nederland

Petra Helmond (Universiteit Utrecht, Pluryn, Universiteit van Amsterdam), Daniel Brugman (Universiteit Utrecht) en Geertjan Overbeek (Universiteit van Amsterdam)

URN:NBN:NL:UI:10-1-100774

Samenvatting

In de huidige studie hebben we de programma-integriteit – de mate waarin een interventie wordt uitgevoerd zoals bedoeld – en effectiviteit van de groepsinterventie EQUIP voor jongeren in justitiële jeugdinrichtingen (jji's) onderzocht. Aan de studie hebben 115 jongeren van zes jji's in Nederland en Vlaanderen deelgenomen. EQUIP bleek effectief in het *neutraliseren* van afnames in sociale vaardigheden en morele waarden die bij de controle groep werden gevonden. Echter, EQUIP bleek niet effectief in het *verbeteren* van sociale vaardigheden en de morele ontwikkeling noch in het *verminderen* van cognitieve vertekeningen. Het EQUIP programma werd met een lage tot matige programma-integriteit uitgevoerd (*M* = 55%). Programma-integriteit had geen modererende werking op de effectiviteit. EQUIP bleek even (in)effectief voor jongeren in de lage als in de matige programma-integriteitsgroep.

Inleiding

Antisociaal gedrag van jongeren wordt breed erkend als een belangrijk maatschappelijk probleem. Antisociaal gedrag kunnen we definiëren als gedrag dat anderen schade toebrengt door het doorbreken van belangrijk sociale, morele of juridische regels (Barriga, Morrison, Liau, & Gibbs, 2001). Antisociaal gedrag bestaat onder andere uit agressieve en delinquente gedragingen zoals mishan-

Correspondentieadres: p.e.helmond@uu.nl / phelmond@pluryn.nl

deling, bedreiging en diefstal. Antisociaal gedrag is niet alleen schadelijk voor slachtoffers, maar ook erg kostbaar voor de maatschappij. In Nederland worden de kosten van criminaliteit geschat op minimaal € 1,239 per hoofd van de bevolking en kost het opsluiten van een jongere in een justitiële jeugdinrichting (jji) € 563 per dag in 2012 (Groot, De Hoop, Houkes, & Sikkel, 2007; Algemene Rekenkamer, 2012).

Er zijn veel interventie programma's ontwikkeld met het doel antisociaal gedrag te verminderen. Uit verschillende meta-analyses blijkt dat met name cognitieve gedragsprogramma's effectief zijn in het verminderen van recidive in vergelijking met gebruikelijke zorg of gedragsprogramma's (Hollin & Palmer, 2009; Landenberger & Lipsey, 2005; Pearson, Lipton, Cleland, & Yee, 2002). In deze effectiviteitsstudies is echter vaak weinig bekend over de manier waarop de interventie in de praktijk wordt uitgevoerd, oftewel de programma-integriteit van een interventie (Durlak & DuPre, 2008; Landenberger & Lipsey, 2005). Programma-integriteit kan worden gedefinieerd als de mate waarin het programma wordt uitgevoerd zoals oorspronkelijk is bedoeld (Caroll et al., 2007; Dane & Schneider, 1998). Een gebrekkig inzicht in de programma-integriteit van een interventie is problematisch, omdat informatie over programma-integriteit inzicht kan geven in waarom een programma wel of niet werkt (Dane & Schneider, 1998; Durlak & DuPre, 2008; Mowbray, Holter, Teague, & Bybee, 2003). Zo kan de afwezigheid van significante interventie effecten enerzijds verklaard worden door een gebrek aan effectiviteit van de interventie zelf, maar kan anderzijds ook worden verklaard door een gebrekkige uitvoering van de interventie.

In dit onderzoek richten we ons op de programma-integriteit van het EQUIP programma voor jongeren in jji's. Het programma beoogt recidive te verminderen door sociale vaardigheden te verbeteren, de morele ontwikkeling te stimuleren en cognitieve vertekeningen te verminderen (Brugman, Bink, Nas, & Van den Bos, 2007). Eerdere studies naar de effectiviteit van EQUIP voor jongeren en volwassenen in een justitiële context laten zowel significante als niet significante effecten zien op cognitieve vertekeningen, sociale vaardigheden, morele ontwikkeling en recidive (Brugman & Bink, 2010; Devlin & Gibbs, 2010; Leeman, Gibbs, & Fuller, 1993; Liau et al., 2004; Nas, Brugman, & Koops, 2005; Van Stam et al., 2014). De EQUIP studies richten zich, zoals de meeste interventie studies in justitiële setting, uitsluitend op het meten van de effectiviteit van het programma. Er is dan ook weinig bekend over de programma-integriteit waarmee EQUIP in de studies werd uitgevoerd en de invloed van de uitvoering op de effectiviteit van EQUIP. Om die reden zullen we in de huidige studie de programma-integriteit *en* effectiviteit van EQUIP voor jongeren in jji's onderzoeken.

Het EQUIP programma

In Nederland is het EQUIP programma in alle jji's in Nederland geïmplementeerd als onderdeel van de basismethodiek YOUTURN (Dienst Justitiële Inrichtingen, 2010). EQUIP is ontwikkeld om antisociale jongeren verantwoordelijk

te leren denken en doen door een combinatie van een wederzijdse hulpbenadering en het aanleren van vaardigheden. De wederzijdse hulpbenadering van EQUIP is gebaseerd op het Positive Peer Culture (PPC) programma (Vorrath & Brendtro, 1985). Het doel van PPC is de negatieve groepscultuur om te buigen naar een positieve groepscultuur waarin jongeren zich verantwoordelijk voor elkaar voelen en elkaar echt helpen (Gibbs et al., 1995). EQUIP gaat uit van de gedachte dat het PPC-model moet worden aangevuld met een programma dat de vaardigheden van antisociale jongeren verbetert (Gibbs et al., 1995). Dit deel van het EQUIP programma is gebaseerd op de Aggression Replacement Training (ART) van Goldstein en Glick (1987). De ontwikkelaars van EQUIP stellen dat antisociale jongeren gekenmerkt worden door drie beperkingen, namelijk (a) een verhoogde mate van cognitieve vertekeningen, (b) tekorten in sociale vaardigheden, en (c) een laag moreel ontwikkelingsniveau.

EQUIP is een multicomponenten programma bestaande uit drie wederzijdse hulp bijeenkomsten en twee equiping bijeenkomsten per week (Gibbs, Potter, & Goldstein, 1995). In EQUIP gebruiken medewerkers en jongeren een gezamenlijke taal van probleemgedragingen en denkfouten (d.w.z., cognitieve vertekeningen) om gedragsproblemen en vertekeningen in het denken te identificeren. In wederzijdse hulp bijeenkomsten leren jongeren elkaar helpen, onder begeleiding van een trainer, bij het identificeren en vervangen van probleemgedragingen en denkfouten. Het equiping curriculum bestaat uit tien woedebeheersingsbijeenkomsten, tien sociale vaardigheidsbijeenkomsten en tien morele keuze bijeenkomsten. In woedebeheersingsbijeenkomsten leren jongeren om hun cognitieve vertekeningen te verbinden aan hun woede en leren ze hoe ze hun woede kunnen controleren en verminderen. In sociale vaardigheidsbijeenkomsten leren jongeren door middel van een stapsgewijze benadering problemen in sociale situaties op te lossen. Tot slot, in morele keuze bijeenkomsten worden jongeren gestimuleerd om moreel volwassen keuzes te maken. In de Nederlandse situatie ten tijde van dit onderzoek werd het EQUIP programma gegeven door de groepsleiders van de leefgroep en/of docenten van het onderwijs.

Effectiviteit van EQUIP

Er zijn vier studies uitgevoerd naar de effectiviteit van EQUIP voor jongeren en volwassenen in een justitiële setting. In de eerste Amerikaanse Randomized Control Trial (RCT) bij mannelijke adolescenten vonden Leeman et al. (1993) dat EQUIP effectief was in het verbeteren van sociale vaardigheden en het verminderen van recidive; EQUIP bleek echter niet effectief in het verbeteren van het moreel redeneren. De Nederlandse quasi-experimentele studie van Nas et al. (2005) bij mannelijke adolescenten liet zien dat EQUIP effectief was in het verminderen van cognitieve vertekeningen, maar niet effectief was in het verbeteren van sociale vaardigheden en moreel redeneren. In een gerelateerde studie bleek EQUIP ook niet effectief in het verminderen van recidive (Brugman & Bink, 2010). In een Amerikaanse RCT studie bij volwassen delinquenten, von-

den Liau et al. (2004) dat EQUIP effectief was in het verminderen van recidive voor vrouwen, maar niet voor mannen. Verder bleek EQUIP niet effectief in het verminderen van cognitieve vertekeningen en het verbeteren van sociale vaardigheden (Liau et al., 2004). Een andere Amerikaanse quasi-experimentele studie bij volwassen delinquenten liet zien dat EQUIP effectief was in het verminderen van recidive voor zowel mannen als vrouwen (Devlin & Gibbs, 2010). Tot slot, laat een recente meta-analyse zien dat EQUIP in een justitiële- en school context een klein tot middelgroot positief effect heeft op sociaal-morele uitkomsten, zoals sociale vaardigheden, morele ontwikkeling en cognitieve vertekeningen (Van Stam et al., 2014). De meta-analyse vindt echter geen significant effect voor de afname van recidive. Amerikaanse studies blijken echter wel een klein tot middelgroot effect te hebben op de afname van recidive, terwijl Europese studies geen effect hebben op recidive. Samenvattend, de beschreven studies laten zien dat EQUIP zowel significante als niet-significante effecten heeft op de doelen van het programma. De verschillen in de effectiviteit van EQUIP kunnen mogelijk verklaard worden door verschillen in kwaliteit van uitvoering van het programma. In de meta-analyse van Van Stam et al. (2014) kon programma integriteit echter niet als moderator worden meegenomen om verschillen in effectiviteit te verklaren, omdat de effectstudies zelden de programma integriteit hebben onderzocht. Onze studie is vernieuwend, omdat wij ons zullen richten op zowel de programma integriteit *als* de effectiviteit van het EQUIP programma in Nederlandse justitiële inrichtingen en hiermee inzicht zullen geven in hoe het programma in de Nederlandse praktijk wordt uitgevoerd.

Programma-integriteit

Onderzoekers onderkennen steeds vaker dat het van cruciaal belang is om programma-integriteit te bestuderen bij de uitvoering van een effectiviteitsstudie van een interventie. Zonder documentatie over programma-integriteit is het moeilijk om te bepalen of significante, niet-significante of ambigue resultaten moeten worden toegeschreven aan het theoretische model waar het programma op berust of aan de uitvoering van het programma (Mowbray et al., 2003). De meerderheid van de effectiviteitsstudies bevat echter geen informatie over de programma-integriteit van de interventie, terwijl studies die *wel* programma-integriteit hebben meegenomen over het algemeen vinden dat een hogere mate van programma-integriteit gerelateerd is aan een hogere mate van programma effectiviteit (Caroll et al., 2007; Durlak & DuPre, 2008; Landenberger & Lipsey, 2005). Volgens Durlak en DuPre (2008) kan de effectiviteit van een interventie zelfs twee tot drie keer groter zijn wanneer het programma integer wordt uitgevoerd.

Ook bij effectiviteitstudies in een justitiële context is vaak weinig bekend over de manier waarop het programma in de praktijk wordt uitgevoerd (Andrews & Dowden, 2005; Landenberger & Lipsey, 2005; Lipsey, 2009). Meta-analyses hebben met behulp van programma-integriteit indicatoren een positieve relatie tussen implementatie kwaliteit en effectiviteit van interventies

gericht op het verminderen van recidive vastgesteld (Andrews & Dowden, 2005; Landenberger & Lipsey, 2005; Lipsey, 2009). Voorbeelden van zulke indicatoren zijn klinische supervisie van staf, trainingshandleidingen en het monitoren van services (Andrews & Dowden, 2005). Om de behandeling van jongeren in een justitiële setting effectiever te kunnen maken is het van belang in kaart te brengen hoe interventies, zoals EQUIP, in de praktijk worden uitgevoerd.

Programma-integriteit wordt als een multidimensioneel concept gezien, bestaande uit blootstelling (exposure), opvolgen van bijeenkomstdoelen (adherence), deelname van participanten (participant responsiveness), en kwaliteit van uitvoering (quality of delivery) (Caroll et al., 2007; Dane & Schneider, 1998). Blootstelling beschrijft in welke mate de voorgeschreven hoeveelheid van het programma wordt uitgevoerd (bijv. de duur en frequentie van de bijeenkomsten), opvolgen van bijeenkomstdoelen geeft aan in hoeverre de inhoud van de bijeenkomsten wordt uitgevoerd zoals bedoeld; deelname van participanten geeft weer in welke mate de deelnemers actief en betrokken zijn tijdens de bijeenkomsten en kwaliteit van uitvoering zegt iets over de mate waarin trainers technieken uit het programma toepassen. De meerderheid van de studies die programma-integriteit hebben onderzocht richt zich slechts op één van de programma-integriteit onderdelen (Durlak & DuPre, 2008). Om een volledig beeld van programma-integriteit te geven, is het echter van belang om alle programma-integriteit onderdelen te onderzoeken. Daarom hebben we in ons instrument alle vier onderdelen van programma-integriteit opgenomen. In onze studie hebben we programma-integriteit gemeten door middel van observaties. Observaties worden als de gouden standaard gezien voor het meten van integriteit (Perepletchikova, 2011). Uit onderzoek blijkt namelijk dat trainers op basis van zelfevaluaties veelal een te positief beeld schetsen over de programma-integriteit waarmee ze de interventie uitvoeren. Programma integriteit op basis van observaties voorspelt de effectiviteit van een programma beter dan programma integriteit op basis van zelf-evaluatie (Durlak & DuPre, 2008; Lillehoj, Griffin, & Spoth, 2004; Vartuli & Rohs, 2009).

De huidige studie

Het doel van de huidige studie is om de programma-integriteit en effectiviteit van EQUIP te onderzoeken door middel van een quasi-experimenteel design met een voor- en nameting in een steekproef van 115 jongeren in jji's in Nederland en Vlaanderen. We verwachten dat jongeren die deelnemen aan EQUIP (d.i. de experimentele groep) een grotere afname in cognitieve vertekeningen en grotere toename van sociale vaardigheden en morele ontwikkeling laten zien dan jongeren die niet deelnemen aan EQUIP (d.i., de controle groep). Verder verwachten we dat programma-integriteit de effectiviteit van EQUIP beïnvloedt: wanneer EQUIP met een hoger niveau van programma-integriteit wordt gegeven zal het programma effectiever zijn in het verminderen van cognitieve vertekeningen en bevorderen van sociale vaardigheden en morele ontwikkeling.

Deze studie kenmerkt zich door een grote klinische relevantie doordat we onderzoeken hoe het EQUIP programma daadwerkelijk wordt uitgevoerd in de jji's. Hierbij worden de verschillende programma-integriteit onderdelen vastgesteld door onafhankelijke observatoren. Verder onderzoeken we hoe effectief EQUIP bij deze uitvoering is in het veranderen van cognitieve vertekeningen, sociale vaardigheden en morele ontwikkeling van jongeren in jji's.

Methode

Steekproef

Jongeren werden geworven bij vijf Nederlandse jji's en één Vlaamse jji. De jongeren verbleven in de jji's wegens een veroordeling van een delict, in afwachting van veroordeling of omdat ze onder toezicht waren geplaatst. Jongeren voor de experimentele groep werden geworven uit eenentwintig EQUIP groepen (zeven meisjesgroepen en veertien jongensgroepen) van de zes deelnemende jji's. In alle jji's werd EQUIP doorlopend gegeven, hetgeen betekent dat jongeren tot een EQUIP groep toetraden en deze verlieten op individuele basis. De EQUIP groepen bestonden meestal uit vijf a zes jongeren. Jongeren uit de controlegroep werden geworven van leefgroepen van twee deelnemende jji's waar EQUIP nog niet was geïmplementeerd. In deze leefgroepen werd het sociaal competentie model gebruikt. Het competentiemodel wordt veelal in jji's gebruikt (Knorth, Klomp, Van den Bergh, & Noom, 2007) en heeft als doel gedragsproblemen te verminderen door de competenties van jongeren te vergroten. Het competentiemodel is met EQUIP samengevoegd tot het YOUTURN programma.

Er deden 234 jongeren mee aan de voormeting van de studie. De definitieve steekproef bestond uit 115 jongeren die zowel een voormeting als nameting hebben in gevuld (n = 89 in de experimentele groep, n = 26 in de controle groep). Een en vijftig procent van de jongeren heeft om diverse redenen geen nameting kunnen invullen, namelijk jongeren werden na rechtbank bezoek vrijgelaten of werden naar een andere jji overgeplaatst of keerden niet terug van verlof. Logistische regressie analyse liet zien dat de conditie, leeftijd, geslacht, etnische achtergrond en voormetingsscores van sociale vaardigheden, moreel redeneren en morele waarden niet gerelateerd waren aan uitval ($p > .05$). Jongeren met minder sterke cognitieve vertekeningen hadden echter een grotere kans om uit te vallen ($p < .05$). De meerderheid van onze definitieve steekproef van 115 jongeren was van het mannelijke geslacht (69%). De gemiddelde leeftijd was 15.54 jaar (SD = 1.56). In ons onderzoek was 61 procent van de jongeren van allochtone afkomst, hetgeen betekent dat minimaal een van de ouders buiten Nederland was geboren. Er werden geen significante verschillen gevonden tussen de experimentele en controle groep met betrekking tot etnische achtergrond, leeftijd, geslacht en voormetingsscores van cognitieve vertekeningen, sociale vaardigheden, moreel redeneren en morele waarden

($p > .05$). De jongeren van de experimentele en controle groep zijn dus door de bank genomen goed vergelijkbaar op de gemeten variabelen.

Procedure

Programma-integriteit

De programma-integriteit werd gemeten door vijf onafhankelijke observatoren: de eerste auteur was getraind in het EQUIP programma en master studenten hadden een twaalfurige observatietraining van de eerste auteur ontvangen. De observatie training bestond uit informatie over het EQUIP programma en het observatie instrument en vier oefen sessies. In iedere EQUIP groep werd één wederzijdse hulp bijeenkomst, één woedebeheersingsbijeenkomst, één sociale vaardigheidsbijeenkomst en één morele keuze bijeenkomst geobserveerd resulterend in een totaal van 83 geobserveerde bijeenkomsten. De interbeoordelaarsbetrouwbaarheid werd bij 23% van de observaties vastgesteld evenredig verdeeld over de type bijeenkomsten. Door de reglementen van de jji's was het verboden om camera's en audiotapes van de bijeenkomsten te maken. Om die reden is programma-integriteit gemeten door middel van directe observatie. Aan trainers en jongeren werd het doel van de observatie uitgelegd en werd de vertrouwelijke context van de bijeenkomst benadrukt. Daarnaast gaven de observatoren aan dat ze niet aan de bijeenkomst deel zouden nemen, maar alleen zouden toekijken.

Programma effectiviteit

Jongeren die aan EQUIP deelnamen werden gevraagd voor en na deelname aan EQUIP vragenlijsten in te vullen met een tijdsinterval tussen de tien en twaalf weken. Jongeren die de jji eerder dan na tien weken verlieten, werden gevraagd de nameting in te vullen indien zij minimaal vier weken aan het programma hadden deelgenomen. De tijdsinterval tussen voor- en nameting was gemiddeld 11.18 weken ($SD = 3.4$). Gezien de variatie in tijdsinterval hebben we deze meegenomen als covariaat in de analyses. Dit tijdsinterval verschilde overigens niet voor de experimentele en controle groep ($p > .05$).

Metingen

Programma-integriteit

De programma-integriteit van EQUIP werd gemeten met de 'Observatie Checklist Programma-integriteit EQUIP'. De checklist werd ontwikkeld op basis van de literatuur over programma-integriteit, op basis van het EQUIP boek en de implementatiegids (Gibbs et al., 1995; Potter, Gibbs, & Goldstein, 2001) en expert consultatie van de ontwikkelaars (J. C. Gibbs, & G. B. Potter, persoonlijke communicaties september en oktober 2008). Het programma-integriteit instrument en trainingsmaterialen kunt u vinden op de http://petrahelmond.com/materials/.

Blootstelling. Het percentage 'blootstelling' werd gemeten aan de hand van drie aspecten: frequentie, duur en uitval van bijeenkomsten. Frequentie van bijeenkomsten geeft het percentage van het aantal bijeenkomsten wat jji's uitvoeren in verhouding tot het aantal bijeenkomsten wat ze zouden moeten uitvoeren volgens het EQUIP boek (Gibbs et al., 1995). De duur van bijeenkomsten geeft het percentage weer van de duur van de bijeenkomsten in verhouding tot de voorgeschreven duur van de bijeenkomsten. De uitval van bijeenkomsten laat het percentage van bijeenkomsten zien wat tijdens de observaties is uitgevallen. De uitval is berekend door het aantal uitgevallen bijeenkomsten te delen door het aantal geplande observaties. Uitval is gehercodeerd zodat bij iedere programma-integriteitsscore een hogere score, een hoger niveau van programma-integriteit weergeeft.

Opvolgen van bijeenkomstdoelen. Het percentage 'opvolgen van bijeenkomstdoelen' geeft weer hoeveel van de inhoudscriteria trainers hebben uitgevoerd tijdens de bijeenkomsten ten opzichte van de inhoudscriteria die uitgevoerd hadden moeten worden tijdens de bijeenkomsten (Gibbs et al., 1995). Gegeven de specifieke inhoud van de verschillende typen bijeenkomsten van het EQUIP programma hebben we aparte observatie formulieren ontwikkeld per type bijeenkomst. Voor wederzijdse hulp, sociale vaardigheden en morele keuze bijeenkomsten is een algemeen formulier ontwikkeld waarin de inhoudscriteria volgens het format van het bijeenkomst type wordt weergegeven. Daarnaast zijn voor sociale vaardigheden en woedebeheersingsbijeenkomsten aparte formulieren ontwikkeld waarin de specifieke inhoud van de tien bijeenkomsten wordt weergegeven. Een voorbeeld item is 'De trainer bespreekt de inhoud van de vorige wederzijdse hulp bijeenkomst' met de categorieën 'Afwezig' (0) of 'Aanwezig' (1). De interbeoordelaarsovereenkomst voor 'opvolgen van bijeenkomstdoelen' was hoog met een gemiddelde Cohen's Kappa van .95 met een range van Kappa .68 tot 1.00.

Deelname van participanten. Het percentage 'deelname van participanten' geeft de geobserveerde deelname van participanten in een EQUIP groep weer ten opzichte van de maximaal haalbare score op deelname. De observatoren scoorden op 19-items om de deelname van participanten tijdens de bijeenkomst te meten. Een voorbeeld item is 'Participanten wijzen groepsleden op hun denkfouten' met antwoord categorieën 'Nooit / zelden' (1) tot 'Meestal / vaak' (4). De interbeoordelaarsovereenkomst voor 'deelname van participanten' was hoog met gemiddelde correlatie van .95.

Kwaliteit van uitvoering. Het percentage 'kwaliteit van uitvoering' geeft de geobserveerde kwaliteit van uitvoering weer ten opzichte van de maximaal haalbare score op kwaliteit. De observatoren gebruikten 16-items om de kwaliteit van uitvoering tijdens de bijeenkomst te scoren. Een voorbeeld items is 'De trainer stimuleerde participanten deel te nemen aan de discussie / mee te denken' met antwoord categorieën 'Nooit / zelden' (1) tot 'Meestal / vaak' (4). De interbeoordelaarsovereenkomst voor 'kwaliteit van uitvoering' was hoog met gemiddelde correlatie van .93.

Programma effectiviteit.

Cognitieve vertekeningen. De schaal cognitieve vertekeningen werd gemeten met de 'Hoe Ik Denk' vragenlijst (HID; Barriga et al., 2001). De HID bevat 39 items betreffende vier categorieën van cognitieve vertekeningen. Participanten geven antwoord op een zespunts Likert schaal 'Erg mee oneens (1) tot 'Erg mee eens' (6). De schaal cognitieve vertekeningen is gebaseerd op het gemiddelde van de 39 items en de betrouwbaarheid was hoog met een Cronbach's alpha van .96 voor de voormeting en .97 voor de nameting.

Sociale vaardigheden. De schaal sociale vaardigheden werd gemeten met de 'Inventory of Adolescent Problems – Short Form Objective' (IAP-SFO), dit is een aangepaste versie van de 'Inventory of Adolescent Problems – Short Form' (Gibbs et al., 1995). In de IAP-SFO worden de sociale vaardigheden van jongeren in een problematische of stressvolle interpersoonlijke relaties getoetst. Jongeren kregen acht situaties met vijf gestandaardiseerde reacties voorgelegd waarbij jongeren de reactie kiezen die het meest op de eigen reactie in de situatie lijkt. De reacties waren ingedeeld op een continuüm van antisociaal via neutraal naar sociaal constructief. De schaal sociale vaardigheden is gebaseerd op de gemiddelde score over de acht situaties en de betrouwbaarheid was hoog met een Cronbach's alpha van .78 voor de voormeting en .82 voor de nameting.

Morele ontwikkeling. De 'Sociomoral Reflection Measure – Short Form Objective' (SRM-SFO) bestaat uit twee schalen, namelijk morele waarden en moreel redeneren (Brugman, Basinger, & Gibbs, 2007). De SRM-SFO bestaat uit tien stellingen met betrekking tot morele waarden op verschillende domeinen. Bijvoorbeeld, 'Hoe belangrijk is het dat mensen zich aan de wet houden?' en 'Waarom is dat belangrijk/niet belangrijk?'. Voor *morele waarden* geven jongeren per stelling aan welk belang zij hechten aan de waarde in een bepaalde stelling met de antwoordcategorieën 'Niet belangrijk' (1) tot 'Erg belangrijk (3). De schaal morele waarden is gebaseerd op het gemiddelde score van de tien stellingen en de betrouwbaarheid was goed met een Cronbach's alpha van .71 voor de voormeting en een Cronbach's alpha voor de .85 voor de nameting.

Vervolgens geven jongeren aan waarom zij de waarde belangrijk vinden door te kiezen uit vier gestandaardiseerde antwoorden die ieder een stadium *moreel redeneren* representeren, zoals beschreven door Gibbs, Basinger, & Fuller (1992). Hierbij geven jongeren eerst aan welke van de redenen dichtbij de eigen reden komt en vervolgens welk antwoord het meest dichtbij de eigen reden komt. Hieruit wordt de score voor de schaal moreel redeneren berekend, welke het stadium van moreel redeneren weergeeft. Voor verdere uitleg over de scoring van de schaal moreel redeneren verwijzen we naar het artikel van Helmond, Overbeek en Brugman (2012) en Beerthuizen (2012). De schaal moreel redeneren was voldoende betrouwbaarheid met een Cronbach's alpha van .61 bij de voormeting en had een goede betrouwbaarheid met een Cronbach's alpha van .85 bij de nameting.

Resultaten

Programma effectiviteit EQUIP

We hebben de effectiviteit van EQUIP getoetst met behulp een MANCOVA met herhaalde metingen (Tabel 1). We vonden significante verschillen tussen de experimentele en controle groep in de ontwikkeling van sociale vaardigheden (p = .016). De experimentele groep bleef stabiel op sociale vaardigheden, terwijl de controle groep een achteruitgang liet zien. Het betreft hier een klein tot middelgroot effect. De experimentele en controle groep verschilden ook significant in morele waarden (p = .035). De experimentele groep bleef stabiel op morele waarden, terwijl de controle groep een afname in het belang van morele waarden liet zien. Ook hierbij ging het om een klein tot matig verschil. We vonden geen verschillen in de experimentele en controle groep in de mate van cognitieve vertekeningen en ontwikkelingsniveau van moreel redeneren (p > .05). De covariaat tijdsinterval tussen voor- en nameting had een significante relatie met cognitieve vertekeningen (p < .030). Voor de controle groep vonden we dat een langere tijdsinterval gerelateerd was aan een toename in cognitieve vertekeningen (p < .044), terwijl we voor de experimentele groep geen significante relatie vonden.

Tabel 1 De effectiviteit van EQUIP op cognitieve vertekeningen, sociale vaardigheden, moreel redeneren en morele waarden

	Experimentele groep				**Controle groep**					
	Voor-meting		**Nameting**		**Voor-meting**		**Nameting**			
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*	η^2_p
Cognitieve vertekeningen	2.56	0.84	2.50	0.89	2.54	1.11	2.48	0.98	0.04	.00
Sociale vaardigheden	0.54	0.83	0.61	0.88	0.92	0.86	0.60	1.10	4.80*	.05
Moreel redeneren	2.91	0.31	2.94	0.34	2.85	0.35	2.88	0.41	0.02	.00
Morele waarden	2.35	0.29	2.33	0.34	2.45	0.30	2.23	0.58	5.00*	.05

Note. Tijdsinterval tussen voor-nameting is meegenomen als covariaat in de analyses
**p < .05 (eenzijgdig)*

Niveaus van programma-integriteit

In Tabel 2 laten we de gemiddeldes, standaarddeviaties en range van programma-integriteit van EQUIP zien (cf. Durlak & DuPre, 2008). De gemiddelde score van de frequentie van bijeenkomsten was 55%, dit betekent dat iets meer dan

de helft van de bijeenkomsten van het EQUIP programma dat is voorgeschreven daadwerkelijk werd uitgevoerd. Het percentage niet-uitgevallen bijeenkomsten was gemiddeld 68%; dit betekent dat een derde van de bijeenkomsten tijdens de observaties uitviel. De duur van de bijeenkomsten had een gemiddelde score van 76%. De bijeenkomsten duurden gemiddeld gezien 46 minuten in plaats van de voorgeschreven 60 minuten. Met betrekking tot het opvolgen van de bijeenkomstdoelen zien we gemiddelde scores van 36% tot 47% voor de verschillende bijeenkomsttypes, dit betekent dat trainers ongeveer een derde tot de helft van de bijeenkomstdoelen opvolgden tijdens de bijeenkomsten. De deelname van de participanten was relatief hoog met 69%, dit is ongeveer twee derde van de hoogst mogelijke score. De kwaliteit van uitvoering had een gemiddelde score van 61%, dit betekent dat trainers ruim over de helft scoorden van de maximaal haalbare score op het gebruik van technieken tijdens het programma.

Tabel 2 Gemiddelde niveaus van programma-integriteit EQUIP (0-100%)

	M	*SD*	*Range*
Samengestelde programma-integriteit	**55%**	**7.25**	**35 – 63%**
Blootstelling	66%	11.85	51 – 85%
Frequentie bijeenkomsten	55%	10.04	50 – 76%
Niet-uitgevallen bijeenkomsten	68%	33.37	0 – 100%
Duur bijeenkomsten	76%	15.28	18 – 88%
Opvolgen bijeenkomstdoelen	43%	10.95	11 – 59%
Wederzijdsehulp	47%	11.15	17 – 67%
Woedebeheersing	40%	14.85	0 – 67%
Sociale vaardigheden	36%	15.97	0 – 71%
Morele keuze	47%	16.15	0 – 71%
Deelname participanten	69%	8.45	47 – 82%
Kwaliteit van uitvoering	61%	6.95	41 – 72%

De modererende rol van programma-integriteit op de effectiviteit van EQUIP

Vervolgens hebben we de modererende rol van programma-integriteit op de effectiviteit van EQUIP onderzocht met behulp van een MANCOVA met herhaalde metingen.

We hebben twee programma-integriteitsgroepen gecreëerd door de experimentele groep op te splitsen op basis van een samengestelde programma-integriteitsscore. De samengestelde programma-integriteitsscore hebben we gecreëerd door het gemiddelde te nemen van de programma-integriteit aspecten. De gemiddelde samengestelde programma-integriteitsscore was 55% met een range van 35% tot 63% (*SD* = 7.3), dit betekent dat gemiddeld gezien onge-

veer de helft van het EQUIP programma werd uitgevoerd. Volgens Durlak en DuPre (2008) kunnen positieve interventie effecten worden verwacht wanneer de programma-integriteit 60% of hoger is. Als we deze drempelwaarde volgen dan kunnen we concluderen dat de programma-integriteit van EQUIP in onze studie laag tot matig was. Daarom hebben we de programma-integriteit groep onder het gemiddelde de lage programma-integriteitsgroep genoemd (n=41) en de groep boven het gemiddelde de matige programma-integriteitsgroep (n=49).

We vonden een significant groepseffect voor de ontwikkeling van sociale vaardigheden (p = .047) (zie Tabel 3). Post-hoc analyse lieten zien dat zowel de lage als matige programma-integriteitsgroepen van de controle groep verschilden op sociale vaardigheden (p = .019; p = .035). De lage en matige programma-integriteitsgroepen bleven stabiel op sociale vaardigheden, terwijl de controle groep een afname in sociale vaardigheden liet zien. De lage en matige programma-integriteitsgroepen verschilden echter niet van elkaar in de ontwikkeling van sociale vaardigheden (p > .05). Ook voor morele waarden vonden we een significant groepseffect (p = .040). We zien dat de lage en matige programma-integriteitsgroepen stabiel bleven op morele waarden en dat de controle groep een afname van morele waarden liet zien (p = .015; p = .033). De lage en matige programma-integriteitsgroepen verschilden echter niet significant van elkaar met betrekking tot morele waarden (p > .05). Tot slot, we vonden geen significante groepsverschillen voor cognitieve vertekeningen en moreel redeneren (p > .05). Wederom had de covariaat tijdsinterval tussen voor- en nameting een significante relatie met cognitieve vertekeningen voor de controle groep (p = .041).

Tabel 3 De modererende rol van programma-integriteit op de effectiviteit van EQUIP

	Experimentele groep								**Controle groep**					
	Lage PI				**Matige PI**									
	Voor-meting		**Na-meting**		**Voor-meting**		**Na-meting**		**Voor-meting**		**Na-meting**			
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*	η^2_p
Cognitieve vertekeningen	2.56	0.80	2.44	0.81	2.56	0.88	2.56	0.97	2.54	1.11	2.48	0.98	0.03	.00
Sociale vaardigheden[a]	0.64	0.83	0.79	0.90	0.45	0.84	0.46	0.84	0.92	0.86	0.60	1.10	2.43*	.05
Moreel redeneren	2.93	0.33	3.01	0.36	2.89	0.29	2.87	0.32	2.85	0.35	2.88	0.41	0.21	.00
Morele waarden[a]	2.33	0.28	2.32	0.34	2.36	0.30	2.34	0.34	2.45	0.30	2.23	0.58	2.60*	.05

Note. Tijdsinterval tussen voor-nameting is meegenomen als covariaat in de analyses; PI = Programma-integriteit
a Lage en Matige PI groepen verschillen significant van de controle groep
**p < .05 (eenzijdig)*

Discussie

In deze studie hebben we de programma-integriteit en effectiviteit van het EQUIP programma voor jongeren in justitiële jeugdinrichtingen (jji's) onderzocht. De studie liet zien dat EQUIP effectief bleek in het *neutraliseren* van afnames in sociale vaardigheden en morele waarden. De EQUIP groep bleef stabiel op sociale vaardigheden en morele waarden, terwijl de controle groep een afname van sociale vaardigheden en morele waarden liet zien. EQUIP bleek echter niet effectief in het *verbeteren* van sociale vaardigheden en de morele ontwikkeling en het *verminderen* van cognitieve vertekeningen. Het EQUIP programma werd met een lage tot matige programma-integriteit uitgevoerd (M = 55%). De resultaten lieten verder zien dat programma-integriteit geen modererende werking had op de effectiviteit van EQUIP. Zowel de lage als de matige programma-integriteitsgroep verschilden van de controle groep in de ontwikkeling van sociale vaardigheden en morele waarden. In tegenstelling tot onze verwachting was EQUIP niet effectiever met een hoger niveau van programma-integriteit. EQUIP bleek even (in)effectief voor jongeren in de lage als in de matige programma-integriteitsgroep.

Uit de studies naar de effectiviteit van EQUIP komt een divers patroon aan uitkomsten naar voren. Net als in de huidige studie, vonden ook Liau et al. (2004) en Nas et al. (2005) dat de EQUIP groep niet vooruit ging in sociale vaardigheden, terwijl Leeman et al. (1993) wel een verbetering van sociale vaardigheden vonden. Ook vonden wij net als Liau et al. (2004) dat EQUIP niet effectief was in het verminderen van cognitieve vertekeningen, terwijl Nas et al. (2005) wel afnames van cognitieve vertekeningen vonden. Tot slot, geen van de studies heeft een effect van EQUIP gevonden op het verbeteren van het moreel redeneren (Leeman et al., 1993; Nas et al., 2005).

In de voorgaande studies was weinig bekend over hoe het EQUIP programma daadwerkelijk werd uitgevoerd. Uit onze studie blijkt nu dat EQUIP met een lage tot matige programma-integriteit niet effectief is in het verminderen van cognitieve vertekeningen, het verbeteren van sociale vaardigheden en de stimulering van morele ontwikkeling. Durlak en DuPre (2008) geven in hun review aan dat positieve interventie effecten bereikt kunnen worden wanneer de programma-integriteit boven de 60% ligt. De programma-integriteit ligt in onze studie echter onder deze drempelwaarde met een gemiddelde van 55% voor de samengestelde programma-integriteit score. Het is daarom wellicht ook niet verwonderlijk dat EQUIP in deze studie niet effectief bleek in het bereiken van de programma doelen.

In tegenstelling tot onze verwachting vonden we dat programma-integriteit de effectiviteit van EQUIP niet modereerde. Onze moderatie hypothese was echter gebaseerd op de verwachting dat de programma-integriteit hoger zou zijn dan in de praktijk het geval was. In onze steekproef bleek echter dat EQUIP alleen werd uitgevoerd met lage tot matige programma-integriteit. De beperkte programma-integriteit range in onze steekproef zou een verklaring kunnen zijn

voor de afwezigheid van de modererende rol van programma-integriteit. Durlak en Dupre (2008) laten in hun review zien dat een gebrek aan variatie in implementatie een verklaring zou kunnen zijn voor een afwezige of zwakke relatie tussen implementatie en effectiviteit in 8 van de 14 studies. "Als de niveaus van implementatie erg hoog of erg laag zijn in de gehele steekproef dan biedt het gebrek aan variatie onvoldoende 'power' om verschillen tussen groepen vast te stellen" (Durlak & Dupre, 2008). Hoewel we geen bewijs vonden voor de modererende rol van programma-integriteit voor de effectiviteit, denken we niet dat dit geïnterpreteerd kan worden als dat het niveau van programma-integriteit irrelevant is voor de effectiviteit van EQUIP. Daartoe ontbreken in deze studie groepen met een voldoende tot hoge programma-integriteit.

Beperkingen

Er zijn verschillende beperkingen aan het huidige onderzoek die benoemd moeten worden. In effectiviteitsstudies heeft een randomized control trial (RCT) de voorkeur boven een quasi-experimenteel design, omdat een RCT potentiele vertekeningen kan elimineren. Het is echter erg moeilijk om een RCT in justitiële context te implementeren (Asscher, Deković, Van der Laan, Prins, & Van Arum, 2007). Ook in onze studie was het niet mogelijk een RCT te implementeren, omdat EQUIP al door verschillende instellingen in de praktijk was geïmplementeerd wat randomisatie binnen een instelling niet mogelijk maakt. Om die reden zijn er buiten de VS, met name ook in Nederland, weinig RCT studies in justitiële context (Asscher et al., 2007; Farrington & Welsh, 2005, Wartna, 2009). Een studie met een hoogstaand quasi-experimenteel design is een goed alternatief om meer te weten te komen over wat werkt in de justitiële setting indien een RCT niet mogelijk is (Hollin, 2008). Een kenmerk van een hoogstaande quasi-experimentele studie is dat de experimentele en controle groep vergelijkbaar zijn op verschillende belangrijke variabelen, zoals ook in onze studie het geval bleek.

Een andere beperking van de huidige studie is het kleine aantal deelnemers aan de controle groep. Tijdens ons onderzoek werd EQUIP in alle jji's in Nederland geïmplementeerd als onderdeel van de basismethodiek YOUTURN (Dienst Justitiële Inrichtingen, 2010). Als direct gevolg van dit beleid was het niet langer mogelijk om jongeren in de controle groep te includeren, omdat alle jongeren in jji's deelnamen aan het EQUIP programma. Een power-analyse liet zien dat we met de huidige steekproef medium effecten kunnen vinden. De kleine steekproef is ook een gevolg van de grote uitval tijdens het onderzoek. De uitval was met name het gevolg van de werkwijze in het Nederlandse justitiële systeem. Uit gegevens van het WODC blijkt dat ongeveer de helft van de jongeren korter dan 3 maanden in de jji verblijft (Repris, 2012). Uit onze uitval analyses bleken weinig verschillen tussen de jongeren die aan de nameting meededen of uitgevallen waren. Echter, jongeren die hoger scoorden op cognitieve vertekeningen hadden een grotere kans om in de studie te blijven, deze jongeren verbleven

vaak lang genoeg om een natmeting in te vullen. Het is daarom van belang dat men voorzichtig is in het generaliseren van de resultaten naar alle jongeren in jji's, omdat de jongeren in onze steekproef een hogere mate van cognitieve vertekeningen lijken te hebben.

Ondanks deze beperkingen kenmerkt de huidige studie zich door een grote klinische relevantie doordat we hebben onderzocht hoe EQUIP daadwerkelijk wordt uitgevoerd in de justitiële context. We hebben hierbij gebruik gemaakt van een multidimensionele meting van programma-integriteit uitgevoerd door onafhankelijke observatoren. Verder hebben we met een quasi-experimenteel voor- en nameting studie onderzocht hoe effectief EQUIP bij deze uitvoering is in het veranderen van cognitieve vertekeningen, sociale vaardigheden en de morele ontwikkeling van jongeren in jji's.

Implicaties

Onze bevindingen hebben verschillende belangrijke implicaties voor onderzoek en praktijk. Al lange tijd zijn er zorgen over de potentiële negatieve effecten van het samen opsluiten van antisociale jongeren in justitiële jeugdinrichtingen (Osgood & Briddell, 2006). Echter, slechts enkele studies hebben deze effecten onderzocht en ondersteunen deze zorgen (Bayer, Pintoff, & Pozen, 2003; Gatti, Tremblay, & Vitaro, 2009; Shapiro, Smith, Malone, & Collaro, 2010). Er zijn verschillende studies die laten zien dat groepsinterventies met antisociale jongeren onbedoelde negatieve effecten kunnen hebben (Dishion & Dodge, 2005; Dishion, McCord, & Poulin, 1999; Poulin, Dishion, & Burraston, 2003), hoewel er ook studies zijn die geen negatieve effecten aantoonden (Handwerk, Field, & Friman, 2000; Weiss et al., 2005). Onze resultaten geven aan dat er in onze studie geen onbedoelde negatieve effecten van de groepsinterventie EQUIP zijn, maar dat het opsluiten in justitiële inrichting mogelijk een negatief effect heeft op de ontwikkeling van sociale vaardigheden en belang van morele waarden. De studie laat zien dat groepsinterventies niet per definitie tot negatieve groepseffecten leiden en dat groepsinterventies mogelijk kunnen helpen bij het neutraliseren van potentiële negatieve groepseffecten in justitiële jeugdinrichtingen.

Er zijn nog twee belangrijk implementatie punten die we willen bespreken die mogelijk een negatieve invloed kunnen hebben gehad op de effectiviteit van EQUIP in onze studie. Het eerste punt betreft de mate waarin het in de Nederlandse justitiële context mogelijk is om een positieve groepscultuur te creëren. Als gevolg van het justitiële systeem in Nederland bestonden de EQUIP groepen niet alleen uit jongeren die veroordeeld waren, maar ook uit jongeren die nog in afwachting van hun uitspraak waren. Sommige jongeren werden na een paar weken vrijgelaten en anderen werden naar een andere instelling overgeplaatst. Hierdoor was het verloop onder de jongeren in de EQUIP groep erg hoog. Dit roept de vraag op of het in deze snel roulerende groepen mogelijk is om een positieve groepscultuur te creëren, gezien het tijd kost om een positieve groepscultuur te ontwikkelen. Het tweede punt betreft de invloed van roule-

rende trainers op de effectiviteit van EQUIP. In het EQUIP programma staat aangegeven dat de bijeenkomsten door een vaste trainer worden uitgevoerd. In strijd met deze aanwijzing werd EQUIP in onze studie (met één uitzondering) door roulerende trainers gegeven. In de Nederlandse situatie werden de bijeenkomsten uitgevoerd door wisselende groepsleiders en docenten. Met frequent wisselende trainers is het de vraag of jongeren een therapeutische relatie met de trainer kunnen ontwikkelen. Het bekend is dat een therapeutische relatie van groot belang is voor de effectiviteit van interventies (Lambert & Barley, 1992). De frequent roulerende jongeren en trainers kunnen mogelijk een negatieve impact hebben gehad op de effectiviteit van EQUIP.

Onze studie heeft een uniek inzicht gegeven in de werkelijke uitvoering van een interventie programma in de justitiële jeugdcontext. We lieten zien dat het EQUIP programma in de dagelijkse praktijk voor een groot deel niet wordt uitgevoerd zoals bedoeld. Naar aanleiding van de studie resultaten hebben we het volgende advies gegeven om de programma integriteit te verbeteren:
1) verhoog de frequentie van bijeenkomsten, met name het aantal wederzijdse hulp bijeenkomsten; 2) verhoog de duur van bijeenkomsten; 3) verminder het aantal bijeenkomsten wat uitvalt; 4) voer meer van de bijeenkomstdoelen uit, naar verwachting is een uitvoering van 60% of meer nodig om effectiviteit te kunnen verwachten (Durlak & DuPre, 2008), 5) gebruik meer van de van de technieken uit het EQUIP programma en verbeter de kwaliteit van uitvoering; 7) gebruik vaste in plaats van roterende trainers; 7) zorg voor centraal management en monitoring van het EQUIP programma (Helmond, Overbeek, & Brugman, 2014). Voor een uitgebreider verbeteradvies zie ons Nederlandse rapport (Helmond, Brugman, & Overbeek, 2009).

Onze bevindingen over de implementatie van EQUIP zijn in overeenstemming met een review waarin implementatie problemen van justitiële interventies in Nederland worden beschreven (Nas, van Ooyen-Houben, & Wieman, 2011). In deze review worden veel implementatie problemen gerapporteerd, zoals het niet uitvoeren van delen van de interventie zoals bedoeld (e.g., niet gebruiken van materialen of handleidingen, inbreng eigen ideeën), gebrek aan steun voor de uitvoering van de interventie, wisselingen in personeel en tekort aan personeel. In een van de weinige empirische studies over kwaliteit van justitiële programma's wordt 68% van de geëvalueerde programma's beoordeeld als ontoereikend (Lowenkamp, Latessa, & Smith, 2006). Deze resultaten geven aan dat de lage tot matige programma-integriteit van EQUIP in Nederlandse jji's niet een uniek geval betreft.

De lage tot matige uitvoering van het EQUIP programma in combinatie met de afwezigheid van positieve interventie uitkomsten onderstrepen het belang van het meten en monitoren van programma-integriteit en effectiviteit van interventies in (justitiële) jeugdzorg. Daarnaast hopen we met deze studie bewustzijn te creëren dat het van groot belang is om interventie programma's met hoge programma-integriteit uit te voeren, zodat 'evidence based' programma's daadwerkelijk effectief kunnen zijn.

Conclusie

EQUIP was effectief in het *neutraliseren* van negatieve effecten op sociale vaardigheden en morele waarden, maar niet effectief in het *verminderen* van cognitieve vertekeningen noch het *verbeteren* sociale vaardigheden en de morele ontwikkeling. Het EQUIP programma werd in de huidige studie met lage tot matige niveaus van programma-integriteit uitgevoerd en programma-integriteit had geen invloed op de effectiviteit van EQUIP. Toekomstig onderzoek moet uitwijzen of EQUIP effectief is wanneer het met hoge niveaus van programma-integriteit wordt uitgevoerd.

Abstract

The present quasi-experimental pre-posttest study examined the program integrity –the extent to which an intervention is implemented as intended– and effectiveness of the cognitive behavioral intervention EQUIP for incarcerated adolescents. Participants (N = 115) were recruited from six correctional facilities in The Netherlands and Flanders. EQUIP was effective in *neutralizing* decreases in social skills and moral value evaluations, but not effective in *reducing* cognitive distortions and *improving* moral judgment. We found low to moderate levels of composite program integrity (M = 55%). Program integrity did not moderate the effectiveness of EQUIP; for both low and moderate program integrity groups EQUIP was equally effective.

Referenties

Algemene Rekenkamer (2012). *Detentie, behandeling en nazorg criminele jeugdigen. Terugblik*. Tweede Kamer, vergaderjaar 2010–2012, 31 215, nrs. 7. Den Haag: Sdu.

Andrews, D. A., & Dowden, C. (2005). Managing correctional treatment for reduced recidivism: A meta-analytic review of programme integrity. *Legal Criminological Psychology, 10,* 173–187.

Asscher, J. J., Deković, M., Van der Laan, P. H., Prins, P. J. M., & Van Arum, S. (2007). Implementing randomized experiments in criminal justice settings: An evaluation of multi-systemic therapy in the Netherlands. *Journal of Experimental Criminology, 3*, 113-129.

Barriga, A. Q., Gibbs, J. C., Potter, G. B., & Liau, A. K. (2001). *How I Think (HIT) questionnaire manual*. Champaign, Illinois: Research Press. (Dutch translation: C. N. Nas (2000). *Hoe Ik Denk Vragenlijst (HID)*. Unpublished manuscript, the University of Utrecht).

Barriga, A. Q., Morrison, E. M., Liau, A. K., & Gibbs, J. C. (2001). Moral cognition: explaining the gender difference in antisocial behaviour. *Merrill-Palmer Quarterly, 47*, 532-562.

Bayer, P., Pintoff, R., & Pozen, D. (2003). *Building criminal capital behind bars: Social learning in juvenile corrections*. Unpublished manuscript, Yale University.

Beerthuizen, M. G. C. J. (2012). The Impact of Morality on Externalizing Behaviour: Values, Reasoning, Cognitive Distortions and Identity (Proefschrift). Retrieved from http://igitur-archive.library.uu.nl.

Brugman, D., Bink, M. D., Nas, C. N., & Van den Bos, J. K. (2007). Kunnen delinquente jongeren elkaar helpen in hun sociale ontwikkeling? Effecten peer-hulpprogramma EQUIP op denkfouten en recidive. *Tijdschrift voor Criminologie, 49*, 153-19.

Brugman, D., Basinger, K. S., & Gibbs, J. C. (2007). *Measuring adolescents' moral judgment: An evaluation of the sociomoral reflection measure – Short Form Objective (SRM-SFO)*. Unpublished manuscript, the University of Utrecht & Urbana University.

Brugman, D., & Bink, M. D. (2010). Effects of the EQUIP peer intervention program on self-serving cognitive distortions and recidivism among delinquent male adolescents. *Psychology, Crime & Law, 17*, 345-358.

Burchinal, M., Xue, Y., Tien, H., Auger, A., & Mashburn, A. (2011). Testing for threshold in associations between child care quality and child outcomes. *Society for Research in Child Development*, Montreal, Canada.

Caroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, *2*(40*)*, 1-9.

Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review, 18*, 23-45.

Devlin, R. S., & Gibbs, J. C. (2010). Responsible Adult Culture (RAC): Cognitive and behavioral changes at a community-based correctional facility. *Journal of Research in Character Education, 8*(1), 1-20.

Dienst Justitiële Inrichtingen Basismethodiek YOUTURN (n.d.). Retrieved August 12, 2010, from http://www.dji.nl/Onderwerpen/Jongeren-in-detentie/Zorg-en-begeleiding/Basismethodiek-YOUTURN/.

Dishion, T. J., & Dodge, K. A. (2005). Peer contagion in interventions for children and adolescents: Moving towards an understanding of the ecology and dynamics of change. *Journal of Abnormal Child Psychology, 33*, 395-400.

Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist, 54*, 755-764.

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41*, 327-350.

Farrington, D. P., & Welsh, B.C. (2005). Randomized experiments in criminology: What have we learned in the last two decades? *Journal of Experimental Criminology, 1*, 9-38.

Gatti, U., Tremblay, R.E., & Vitaro, F. (2009). Iatrogenic effect of juvenile justice. *Journal of Child Psychology and Psychiatry, 50*, 991–998.

Gibbs, J. C., Basinger, K. S., & Fuller, D. (1992). *Moral maturity: Measuring the development of sociomoral reflection*. Hillsdale, NJ: Erlbaum.

Gibbs, J. C., Potter, G. B., & Goldstein, A. P. (1995). *The EQUIP Program: Teaching youth to think and act responsibly through a peer-helping approach*. Champaign, IL: Research Press.

Goldstein, A. P., & Glick, B. (1987). *Aggression Replacement Training: A comprehensive interventions of aggressive youth*. Champaign, IL: Research Press.

Groot, I., De Hoop, T., Houkes, A., & Sikkel, D. (2007). *De kosten van criminaliteit. Een onderzoek naar de kosten van criminaliteit voor tien verschillende delicttypen*. Amsterdam: SEO.

Handwerk, M. L., Field, C. E., & Friman, P. C. (2000). The iatrogenic effects of group intervention for antisocial youth: Premature extrapolations? *Journal of Behavioral Education, 10*, 223-238.

Helmond, P. E., Brugman, D., & Overbeek, G. (2009). Programma integriteit EQUIP residentiele inrichtingen in Nederland en België. Unpublished document, Department of Developmental Psychology, Utrecht University, Utrecht.

**Helmond, P. E., Overbeek, G., & Brugman, D. (2012). Program integrity and effectiveness of a cognitive behavioral intervention for incarcerated youth on cognitive distortions, social skills, and moral development. *Children and Youth Services Review, 34*, 1720- 1728.

Helmond, P., Overbeek, G., & Brugman, D. (2014). Boosting program integrity and effectiveness of the cognitive behavioral program EQUIP for incarcerated youth in The Netherlands. *Children and Youth Services Review*, 39, 108-116 .

Hollin, C. R. (2008). Evaluating offending behaviour programmes: Does only randomization glister? *Criminology and Criminal Justice, 8,* 89-106.

Hollin, C. R., & Palmer, E.J. (2009). Cognitive skills programmes for offenders. *Psychology, Crime & Law, 15*, 147-164.

Knorth, E. J., Klomp, M., Van den Bergh, P. M., & Noom, M. J. (2007). Aggressive adolescents in residential care: A selective review of treatment requirements and models. *Adolescence, 42,* 461-486.

Lambert, M. J., & Barley, D. E. (2001). Research summary on the therapeutic relationship andpsychotherapy outcome. *Psychotherapy: Theory, Research, Practice, Training, 38*, 357-361.

Landenberger, N. A., & Lipsey, M. W. (2005). The positive effects of cognitive-behavioral programs for offenders: A meta-analysis of factors associated with effective treatment. *Journal of Experimental Criminology, 1*, 451-476.

Leeman, L. W., Gibbs, J. C., & Fuller, D. (1993). Evaluation of a multi-component group treatment program for delinquents. *Aggressive Behaviour, 19*, 281-292.

Liau, A. K., Shively, R., Horn, M., Landau, J., Barriga, A., & Gibbs, J. C. (2004). Effects of psychoeducation for offenders in a community correctional facility. *Journal of Community Psychology, 32*, 543-558.

Lillehoj, C. J. G., Griffin, K. W., & Spoth, R. (2004). Program provider and observer ratings of school-based preventive intervention implementation: Agreement and relation to youth outcomes. *Health Education & Behavior, 31*, 242-257.

Lipsey, M. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims & Offenders, 4*, 124-147.

Lowenkamp, C. T., Latessa, E. J., & Smith, P. (2006). Does correctional program quality really matter? The impact of adhering to the principles of effective intervention. *Criminology and Public Policy, 5*, 575-594.

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Developmental, measurement, and validation. *American Journal of Evaluation, 24*, 315–340.

Nas, C. N., Brugman, D., & Koops, W. (2005). Effects of a multi-component peer intervention for juvenile delinquents on moral judgment, cognitive distortions, and social skills. *Psychology, Crime & Law, 11*, 421-434.

Nas, C. N., Van Ooyen-Houben, M. M. J., & Wieman, J. (2011). *Interventies in uitvoering. Wat er mis kan gaan bij de uitvoering van justitiële (gedrags)interventies en hoe dat komt.* Den Haag: WODC.

Osgood, D. W., & Briddell, L. O. (2006). Peer effects in juvenile justice. In K. A. Dodge, T. J. Dishion, & J. E. Lansford (Eds.), *Deviant peer influences in programs for youth* (pp. 141–161). New York: Guilford.

Pearson, F. S., Lipton, D. S., Cleland, C. M., & Yee, D. S. (2002). The effects ofbehavioral/cognitive-behavioral programs on recidivism. *Crime and Delinquency, 48*, 476–496.

Perepletchikova, F. (2011). On the topic of treatment integrity. *Clinical Psychology: Science and Practice, 18,* 148-153.

Potter, G. B., Gibbs, J. C., & Goldstein, A. P. (2001). *EQUIP implementation guide*. Champaign, IL: Research Press.

Poulin, F., Dishion, T. J., & Burraston, B. (2001). 3-Year iatrogenic effects associated with aggregating high-risk adolescents in cognitive–behavioral preventive interventions. *Applied Developmental Science, 5*, 214-224.

Repris (2012, October, 10). *Repris WODC-recidive monitor - verblijfsduur ex-JJI pupillenuitstroomjaren 2006-2008*. Retrieved from http://www.wodc.nl/onderzoek/ cijfers-en-prognoses/ Recidive-monitor/Repris/index.aspx

Shapiro, C. J., Smith, B. H., Malone, P. S., & Collaro, A. L. (2010). Natural Experiment in Deviant Peer Exposure and Youth Recidivism. *Journal of Clinical Child & Adolescent Psychology, 39*, 242-251.

Spoth, R., Guyll, M., Trudeau, L., & Goldberg-Lillehoj, C. (2002). Two studies of proximal outcomes and implementation quality of universal preventive interventions in a community-university collaboration context. *Journal of Community Psychology, 30*, 499-518.

Van Stam, M. A., Van der Schuur, W. A., Tserkezis, S., Van Vugt, E. S., Asscher, J. J., Gibbs, J. C., & Stams, G. J. J. M. (2014). The Effectiveness of EQUIP on Sociomoral Development and Recidivism Reduction: A Meta-Analytic study. *Children and Youth Services Review, 38*, 44–51.

Vartuli, S., & Rohs, J. (2009). Assurance of outcome evaluation: Curriculum fidelity. *Journal of Research in Childhood Education, 23*, 502-512.

Vorrath, H. H., & Brendtro, L. K. (1985). *Positive peer culture* (2nd ed.). Chicago: Aldine.

Wartna, B. S. J. (2009). *In de oude fout.* Wetenschappelijk Onderzoek- en Documentatiecentrum.

Weiss, B., Caron, A., Ball, S., Tapp, J., Johnson, M., & Weisz, J. (2005). Iatrogenic effects of group treatment for antisocial youth. *Journal of Consulting and Clinical Psychology, 73*, 1036-1044