# UvA-DARE (Digital Academic Repository)

## An examination of program integrity and recidivism of a cognitive-behavioral program for incarcerated youth in The Netherlands

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

[Link to publication](Link to publication)

# An examination of program integrity and recidivism of a cognitive-behavioral program for incarcerated youth in The Netherlands

Petra Helmond[a,b,c]*, Geertjan Overbeek[c] and Daniel Brugman[a]

[a]Department of Developmental Psychology, Utrecht University, Utrecht, The Netherlands;
[b]Department Research & Development, Pluryn, Nijmegen, The Netherlands; [c]Research Institute of Child Development, University of Amsterdam, Amsterdam, The Netherlands

The present study examined whether the cognitive behavioral program EQUIP for incarcerated youth would reduce recidivism and whether higher levels of program integrity – the extent to which a program is implemented as intended – would strengthen the effectiveness of EQUIP on recidivism. Program integrity was measured using a multifaceted program integrity instrument assessing the level of exposure, adherence, participant responsiveness, and quality of delivery. Participants ($N = 133$) were recruited from five juvenile correctional facilities in the Netherlands. The EQUIP program was implemented with low to moderate levels of program integrity ($M = 54\%$). No differences were found between the experimental and control group in the prevalence, frequency, and seriousness of recidivism. In addition, within the experimental group, program integrity did not strengthen the effectiveness of EQUIP on the prevalence, frequency, and seriousness of recidivism. With these low-to-moderate levels of program integrity EQUIP did not reduce recidivism.

**Keywords:** effectiveness; recidivism; program integrity; EQUIP; delinquency; adolescents

Correctional treatment researchers have written extensively about the importance of program integrity of rehabilitation programs (Andrews & Dowden, 2005; Gendreau, Goggin, & Smith, 1999; Landenberger & Lipsey, 2005; Lipsey, 2009). Many intervention studies, especially those conducted in the field of correctional treatment, have failed to include measures of program integrity on the actual implementation of an intervention (Andrews & Dowden, 2005; Durlak & DuPre, 2008; Landenberger & Lipsey, 2005). Only 3.5% of intervention studies published in high-quality clinical journals adequately assessed program integrity (Perepletchikova, Treat, & Kazdin, 2007). This is highly problematic because without information on program integrity it is unclear whether the positive, negative, or absent intervention effects can and should be attributed to the intervention or to a failure to implement the program as intended. In the absence of measurements of program integrity in most correctional treatment studies (Andrews & Dowden, 2005; Landenberger & Lipsey, 2005; Lipsey, 2009), meta-analyses used proxies of program integrity to establish its relation with recidivism. Examples of these proxies are clinical supervision by staff, presence of training manuals, monitoring of service

---

process, and adequate dosage (Andrews & Dowden, 2005). With these program integrity proxies meta-analyses have established very global, but positive, relations between program integrity and effectiveness of programs aimed at reducing recidivism (Andrews & Dowden, 2005; Landenberger & Lipsey, 2005; Lipsey, 2009). In addition, a few empirical studies showed that program integrity, defined as the adherence to effective principles of correctional treatment, is related to reductions in recidivism (Lowenkamp, Latessa, & Smith, 2006; Lowenkamp, Makarios, Latessa, Lemke, & Smith, 2010). In the present study we will focus on the program integrity and recidivism of the cognitive behavioral program EQUIP (Gibbs, Potter, & Goldstein, 1995). In our study program, integrity is defined as the extent to which a program is actually implemented as designed (Caroll et al., 2007; Dane & Schneider, 1998). Previous studies on the effectiveness of EQUIP showed inconsistent results (Brugman & Bink, 2011; Devlin & Gibbs, 2010; Leeman, Gibbs, & Fuller, 1993; Liau et al., 2004; Nas, Brugman, & Koops, 2005). Because these studies did not include measures on program integrity, it is unclear whether the program was implemented as intended in these studies. The aim of the present study is to investigate the effectiveness of EQUIP in reducing recidivism and to examine whether EQUIP is more effective in reducing recidivism when it is delivered with higher levels of program integrity.

### The EQUIP program

EQUIP is a cognitive behavioral program designed to teach incarcerated youth to think and to act responsibly by combining a peer helping and a skills streaming approach. The peer helping approach of the EQUIP program is based on the Positive Peer Culture (PPC) Model (Vorrath & Brendtro, 1985). The PPC Model aims to transform a negative peer culture into a positive one, in which individuals feel responsible for each other and help one another (Gibbs et al., 1995). However, a peer helping approach alone is not sufficient to counter negative peer pressure, since antisocial youth often lack the skills necessary to adequately help each other (Gibbs et al., 1995). The EQUIP program therefore also targets three specific 'limitations' of antisocial youth: cognitive distortions, social skill deficiencies, and moral developmental delays (Helmond, Overbeek, & Brugman, 2012; Gibbs et al., 1995). These limitations are addressed in the skills streaming curriculum of EQUIP that is based on Aggression Replacement Training (ART; Glick, & Gibbs, 2011; Goldstein & Glick, 1987). The first limitation, cognitive distortions, can be described as 'inaccurate or rationalizing attitudes, thoughts or beliefs concerning one's own or other's behavior' (Gibbs et al., 1995, p. 108). The second limitation, social skills deficiencies, is defined as 'imbalanced and unconstructive behavior in difficult interpersonal situations' (Gibbs et al., 1995, p. 165). The third limitation, moral developmental delays, can be defined as 'the persistence beyond early childhood of an immature moral judgment and a pronounced "me-centeredness" or egocentric bias' (Gibbs et al., 1995, p. 43). Many previous studies have shown that cognitive distortions, poor social skills, and immature moral judgments are related to antisocial behavior (Barriga, Hawkins, & Camelia, 2008; Beauchamp & Anderson, 2010; Helmond, Overbeek, Brugman, & Gibbs, 2014; Lösel & Beelmann, 2003; Nas, Brugman, & Koops, 2008; Raaijmakers, Engels, & Van Hoof, 2005; Stams et al., 2006; Van Vugt et al., 2011). The EQUIP program targets several important dynamic needs, especially the big four identified by Andrews, Bonta, and Wormith (2006), such as problem-solving skills, anger management, reduction of

antisocial cognition, recognition and reduction of risky thinking, and reduction of antisocial associates.

In the EQUIP program, staff and youth use a common program language of problem names and thinking errors (i.e., cognitive distortions) to identify behavioral problems and distorted thinking. EQUIP consists of both mutual help meetings, and equipment meetings. In mutual help meetings youths work on identifying and replacing problem names and thinking errors with the help of their group under guidance of a trainer. The multicomponent equipment meetings consist of 10 anger management meetings, 10 social skills training meetings, and 10 social decision-making meetings. In anger management and thinking error correction meetings, youths learn to connect (distorted) thinking to anger and how to control and reduce their anger. In social skills meetings, youths learn to solve problems in social situations in a step-by-step approach. Finally, in social decision-making meetings youths are facilitated in making more mature moral judgments. EQUIP groups are supposed to meet for minimally three mutual help meetings and two equipment meetings a week (Gibbs et al., 1995). The equipment curriculum can thus be completed in 10 weeks, when splitting up the social skills training across the two equipment meetings and combining it with anger management and social decision-making meetings (Gibbs et al., 1995). Each meeting lasts one to one and a half hours. Group meetings are 'sacred'; therefore, cancelation of meetings should be prevented at all times.

## The effectiveness of EQUIP

Until now, six studies have been published on the effectiveness of EQUIP for incarcerated offenders. These studies showed both significant and nonsignificant effects on the targeted dimensions of the EQUIP program. Some studies showed effects on the increase of social skills (Leeman et al., 1993), the reduction of cognitive distortions (Brugman & Bink, 2011; Nas et al., 2005), and the reduction of recidivism (Devlin & Gibbs, 2010; Leeman et al., 1993; Liau et al., 2004). Other studies, however, did not find significant effects on moral reasoning (Leeman et al., 1993; Nas et al., 2005), social skills (Liau et al., 2004; Nas et al., 2005), cognitive distortions (Liau et al., 2004), or recidivism (Brugman & Bink, 2011; Liau et al., 2004). Previous studies on EQUIP did not take measures of program integrity into account. Consequently, little is known about the actual implementation of the EQUIP program at the time of these studies. Nas et al. (2005) and Brugman and Bink (2011) reported concerns on a weaker implementation of EQUIP, specifically the absence of mutual help meetings and a positive peer culture. Also Liau et al. (2004) reported a lower frequency of meetings than prescribed. In the present study, we focus on program integrity of EQUIP as a potential factor for explaining differences in outcomes. In a recent quasi-experimental study on the effectiveness of EQUIP, we included measures of program integrity (Helmond et al., 2012). In that study, we examined the effectiveness of EQUIP on process outcomes, i.e., the underlying social cognitive processes that EQUIP targets to promote behavioral change. In that study, we showed that levels of program integrity of EQUIP in juvenile correctional facilities in the Netherlands and Flanders were low-to-moderate ($M = 55\%$). With these low-to-moderate levels of program integrity, the EQUIP program did not show the expected intervention effects. Both the EQUIP and the control group remained stable on cognitive distortions and moral judgment. However, youths receiving EQUIP did remained stable in social skills and moral values, whereas their peers in a control group showed a decrease in

social skills and moral values. In a related study, with a new EQUIP sample of incarcerated youths, we found similar levels of program integrity and again we did not find the expected improvements on youth's process outcomes (Helmond, Overbeek, & Brugman, 2014). As a next step, the present study focuses on whether EQUIP is effective on behavioral outcomes, i.e., in reducing the likelihood of recidivism, and whether higher levels of program integrity strengthens the effectiveness of EQUIP on recidivism.

## Program integrity

One of the few empirical assessments of program integrity in correctional treatment can be found in studies using the Correctional Program Assessment Inventory (CPAI; Lowenkamp, Latessa, & Smith, 2006; Lowenkamp, Makarios, et al., 2010). These studies showed that higher levels of program integrity were related to greater reductions of recidivism. A condensed survey-based version of the CPAI, however, did not relate to recidivism of incarcerated youth (Sullivan & Latessa, 2011). The CPAI focuses on organizational features that are essential for proper delivery of a correctional treatment or so-called 'effective principles' of correctional treatment, such as program and staff characteristics. As such the CPAI does not tap into the actual implementation of a specific correctional program. Barnoski (2004) demonstrated that Family Functional Therapy (FFT) and Aggression Replacement Training (ART) produced greater reductions in recidivism when implemented competently. However, a major shortcoming of this study was that the measurement of 'competence' was based on post-hoc recollections of involved supervising staff rather than on real-time measurement. To overcome this 'program integrity' gap in correctional treatment literature, the present study provides a thorough assessment of program integrity of a specific correctional program and will investigate whether program integrity can predict outcomes on recidivism. In contrast to CPAI, our program integrity assessment focuses on the internal aspects of a specific program, including the direct face-to-face interaction between program staff and offenders (McGuire, 2000).

Because no instrument existed yet to assess the program integrity of EQUIP, we designed such an instrument (Helmond, Overbeek, & Brugman, 2013). Program integrity is described to be a multifaceted construct and has often been described to include the following facets: exposure, adherence, participant responsiveness, and quality of delivery (Caroll et al., 2007; Dane & Schneider, 1998). Exposure describes the length and frequency of the sessions implemented by the facility; adherence refers to the degree to which meetings are delivered as prescribed; participant responsiveness gives insight into the degree to which participants are engaged and involved in the meetings; and quality of delivery describes the manner in which trainers use the techniques and methods as prescribed. Even though program integrity is acknowledged to be multifaceted, the majority of empirical studies that included program integrity instruments tapped only into one of the facets (Durlak & DuPre, 2008). To fully account for the different facets of program integrity, however, it is crucial to include all four in its measurement. In addition, in our study program integrity will be assessed by independent observers and not by trainer's self-evaluations. Observations are viewed as the most robust measurement of integrity (Allen, Linnan, & Emmons, 2012; Perepletchikova, 2011). Observations are seen as a more realistic assessment, as trainers' self-evaluations tend to be positively biased (Durlak & DuPre, 2008; Lillehoj, Griffin, & Spoth, 2004; Vartuli & Rohs, 2009). In addition, there are indications that program integrity assessed by observers is more

often related to program effectiveness than self-evaluations (Durlak & DuPre, 2008; Lillehoj et al., 2004; Vartuli & Rohs, 2009).

### The present study

The aim of the present study was to examine the effectiveness of EQUIP on recidivism in a sample of 133 youths incarcerated in correctional facilities in the Netherlands. We investigated whether youths participating in EQUIP (i.e., the experimental group) showed a lower prevalence, frequency, and seriousness of recidivism compared with youths not participating in EQUIP (i.e., the control group). In addition, we hypothesized that higher levels of program integrity of EQUIP were related to a lower prevalence, frequency, and seriousness of recidivism.

## Method

### Sample

In the present quasi-experimental study participants were recruited from five comparable high-security Dutch juvenile correctional facilities. The participants were incarcerated for committing crimes, were awaiting sentencing, or were placed under supervision order. Participants in the experimental condition were recruited from 19 EQUIP groups (7 female and 12 male EQUIP groups). Participants in the control condition were recruited from units of two of the participating correctional facilities in which EQUIP had not been implemented yet. In these units the Social Competence Model (SCM) was used. SCM is aimed at reducing externalizing problem behavior and increasing social competencies of juveniles. SCM is a frequently used method in Dutch juvenile correctional facilities, thus representing usual care in the Netherlands (Knorth, Klomp, Van den Bergh, & Noom, 2007).

Sixty-three percent of the participants who completed a pretest dropped out of the study for several reasons: participants were released after court visit, were transferred to a different facility, and a few did not return from furlough. The sample was further reduced, because 13 participants had not yet been released at the time of measurement of recidivism and 10 official records could not be traced. The final sample consisted of 133 participants with $n = 110$ in the experimental group and $n = 23$ in the control group. A logistic regression analysis showed that in the experimental condition, age, gender, ethnic background, and pretest scores of social skills, moral judgment and moral value evaluation were all unrelated to attrition, respectively (OR = 1.041, $p = .912$; OR = 1.181, $p = .063$; OR = .984, $p = .964$; OR = 1.391, $p = .197$; OR = .896, $p = .580$; OR = .995, $p = .191$; OR = .814, $p = .634$). However, participants with less severe cognitive distortions at pretest were more likely to drop out of the sample from pre- to posttest (OR = .543, $p = .002$).

Table 1 presents the descriptive of the final sample. The majority of our final sample of 133 participants was boys (74%), and the mean age at pretest was 15.7 years (SD = 1.5). In this study, 59% of the participants had an ethnic minority status, meaning that at least one of the youth's parents was born outside the Netherlands. No differences were found between the experimental and control groups concerning age, ethnic minority status, criminal law placement (*vs.* placement under supervision order), age of first offense, frequency of previous offenses, duration of stay for current offense, observation period of recidivism, respectively ($F$ (1, 131) = .42, $p = .517$; $\chi^2$ (1) = .48, $p = .490$); $\chi^2$ (1) = 1.03, $p = .309$;

Table 1. Sample characteristics of final sample with available recidivism data.

|  | Min–Max | Total | Control group | Experimental group |
|---|---|---|---|---|
| Boys | 0–1 | 74% | 52% | 78%* |
| Ethnic minority | 0–1 | 59% | 65% | 57% |
| Criminal law placement | 0–1 | 44% | 35% | 46% |
| Age | 12–18 | 15.7 (1.5) | 15.8 (1.6) | 15.6 (1.4) |
| Age first crime | 12–18 | 14.2 (1.5) | 14.6 (1.7) | 14.1 (1.4) |
| Frequency previous offenses | 0–26 | 3.8 (4.6) | 2.7 (3.5) | 4.1 (4.8) |
| Duration of stay (months) | 1.6–39.6 | 6.2 (4.5) | 6.1 (4.8) | 5.8 (3.0) |
| Observation period (months) | 3.4–57.8 | 18.7 (5.0) | 17.3 (2.3) | 19 (5.3) |
| Seriousness previous offenses (0 = no previous offense) | 0–3 | 1.9 (1.2) | 1.4 (1.3) | 2.0 (1.2)* |

*$p < .05$.

$F(1, 128) = 1.22, p = .237$; $F(1, 128) = 1.56, p = .214$; $F(1, 130) = .21, p = .646$; $F(1, 130) = 2.344, p = .128$). However, we did find differences between the experimental and control group in gender distribution and seriousness of previous offenses ($\chi^2 (1) = 6.64, p = .010$; $F(1, 130) = 4.06, p = .046$). The experimental group included more boys and youths in the experimental group had committed more severe previous offenses when compared with the control group. Consequently, gender and seriousness of previous offenses were included as covariates in the analyses.

## Measures

### Recidivism

To establish whether the participants had reoffended since their release from the institution official records were requested from the Judicial Information Service (JustID). In addition, data on entry and release dates of the youth were obtained from Custodial Institutions Agency (DJI). The official records were coded using the Recidivism Coding System (RCS) of Research and Documentation Centre (WODC) of the Ministry of Justice (Wartna, Blom, & Tollenaar, 2011; Wartna, El Harbachi, & Van der Laan, 2005). In accordance with the RCS guidelines, minor offenses like traffic offenses were not taken into consideration. In line with RCS, offenses were included if they were classified having a 'valid disposal,' meaning that cases were settled by the public prosecutor by means of a discretionary dismissal or a transaction in which the judge gives a guilty verdict (Wartna et al., 2005, 2011). Following the RCS guidelines, cases that have not yet been settled or that are being heard on appeal were also included as recidivism, as 9 out of 10 cases ends up classified having a valid disposal (Wartna et al., 2005, 2011). Furthermore, we used the RCS to code the seriousness of offenses into minor offenses, serious offenses, and very serious offenses (Wartna et al., 2005, 2011). The observation period for the measurement of recidivism started at the moment the youngsters were released from the institution and ended on the day that the official records were released by JustID.

We included three types of recidivism measures: prevalence, frequency, and seriousness of recidivism. 'Prevalence of recidivism' was coded as 'recidivism' *(1)*, i.e., a youth reoffended after release or as 'no recidivism' *(0)*, i.e., a youth did not reoffend after release. 'Frequency of recidivism' was coded as the number of repeated offenses after release. The 'seriousness of recidivism' was coded as 'no offenses' *(0)*, minor offenses *(1)*, 'serious

offenses' *(2)*, and 'very serious offenses' *(3)*. Examples of minor offenses were slight molestation, vandalism, nonviolent property offense, and examples of serious offenses were swindle, theft, and burglary, and examples of very serious offenses were manslaughter, rape, and grievous bodily harm (Wartna et al., 2005, 2011).

### Program integrity

Program integrity was measured by nine trained independent observers. The observation training consisted of information on the EQUIP program, the observation instrument, and four practice sessions. In each EQUIP group we observed one mutual help meeting, one anger management meeting, one social skills training meeting, and one social decision-making meeting for each EQUIP group. Due to the correctional facility regulations, cameras or audio-tapes to record meetings were forbidden. Consequently, we assessed program integrity by direct observations. Trainers were informed about the purpose of the observations and when observations were scheduled. The observers explained the purpose of their presence to the EQUIP group and stressed the confidential nature of the observations and also explained that they would not participate in the meeting.

The program integrity of EQUIP was measured using the 'Observation Checklist Program Integrity EQUIP.' The observation checklist includes the four facets of program integrity: exposure, adherence, participant responsiveness, and quality of delivery (Caroll et al., 2007; Dane & Schneider, 1998; Durlak & DuPre, 2008; Mowbray, Holter, Teague, & Bybee, 2003). Content of the measures was based on the EQUIP book and implementation guide (Gibbs et al., 1995; Potter, Gibbs, & Goldstein, 2001) and expert consultations with the intervention's authors (J. C. Gibbs, & G. B. Potter, personal communication, 4 September 2008, 9 September 2008, 9 October 2008). Specific information on the observation checklist can be requested from the first author. The 'Observation Checklist' as part of the Measurement Instrument Program Integrity EQUIP (MIPIE) showed good psychometric quality, in terms of construct validity, internal consistency of the composite scale, and inter-observer agreement (Helmond et al., 2013). Perepletchikova (2011) describes a continuum on the adequacy of program integrity procedures. Based on Perepletchikova's (2011) continuum on the adequacy of program integrity procedures, our instrument is at the recommended level of rigor for RCTs.

### Exposure

The frequency of meetings is the percentage of the program meetings acquired by dividing the number of meetings that institutions intended to implement over a 10-week period by the number of meetings that should have been implemented during this period according to the EQUIP program (Gibbs et al., 1995). The cancelation of meetings reflects the percentage of meetings canceled as determined during the observation of meetings. The cancelation percentage is calculated by dividing the number of canceled meetings during the observations by the number of scheduled observation meetings. The percentage of canceled meetings was reversely coded into uncanceled meetings so that a higher program integrity score indicates a higher level of program integrity for all program integrity aspects. The duration time of the meetings reflects the percentage of effective meeting time relative to the prescribed minimum meeting time (i.e., 60 minutes).

*Adherence*

Adherence refers to the percentage of content criteria attained during the meeting divided by the number of content criteria that should have been present during the meeting according to the EQUIP program (Gibbs et al., 1995). Given the specific content of each EQUIP meeting type, we developed separate observation forms for each of the meetings. For mutual help, social skills, and social decision-making meetings, a general form reflecting the format of the meeting type was developed. In addition, for the social skills and anger management meetings, specific forms were developed reflecting the specific content of each of the 10 meetings. An example item is 'The trainer reviews the content of the previous mutual help meeting' with categories absent (0) or present (1).

*Participant responsiveness*

This measure reflects the observed responsiveness of all participants in an EQUIP group relative to the highest possible responsiveness rate. Observers scored 19 items to assess the participants' responsiveness during the meeting. Two example items are 'Participants are negative: resistant, sullen, do not want to be there' with categories 'Characteristic for none (1) to all (5) of the participants,' and 'Participants point out other group members' thinking errors' with answer categories never/seldom (1) to most of the time/often (4). The presented answer categories were used for most items.

*Quality of delivery*

Observers rated the quality of delivery on a 16 item scoring card developed to assess the trainer's use of required techniques during the meeting. An example item of the questionnaire is 'The trainer encourages participants to participate in discussion/thinking along' with answer categories never/seldom (1) to most of the time/often (4). These answer categories were used for most items.

*Composite program integrity*

We created a composite program integrity score by taking the average of the program integrity aspects, i.e., frequency of meetings, cancelation of meetings, meeting time, adherence to mutual help, anger management, social skills, and social decision-making meetings, quality of delivery, and participant responsiveness.

## Strategy of analysis

Our effectiveness data has a multilevel structure with participants (Level One) nested in treatment groups (Level Two). A well-known problem of ignoring dependency in multilevel data by using one-level instead of two-level models is that the significance level of the findings may be biased (Hox, 2010). Therefore, we tested whether our data had a multilevel structure of our recidivism outcomes in MLwiN 2.21 (Rasbash, Charlton, Browne, Healy, & Cameron, 2010). We found that the two-level model did not have a significantly better fit compared with the one-level model for the prevalence, survival time, frequency, and seriousness of recidivism (for all variables: –2LL deviance: 0.000, $p = .50$). Therefore, we continued our analyses in a one-level model in SPSS.

We used survival analysis to analyze the effectiveness of EQUIP on the prevalence of recidivism. Survival analysis involves the modeling of time to event (i.e., recidivism) data and takes censoring into consideration (Kleinbaum & Klein, 2005). In our sample, there

was variation in the length of the observation period after release because participants had left the facility at different dates. Censored cases are participants that did non-recidivate during the observation period. We used Cox regression to examine differences in the prevalence of recidivism between experimental and control group. Given the significant differences between the experimental and control group in gender distribution and seriousness of previous offenses, these variables were included as covariates in the Cox regression. Survival analyses are performed on dichotomous dependent variables. Therefore, we performed, respectively, Poisson and hierarchical regression analyses to examine group differences in the frequency and seriousness of recidivism again including gender and seriousness of previous offenses as covariates. In Poisson regression models, it often occurs that the model is overdispersed, meaning that the variance of the dependent variable is larger than the mean. We corrected for this by including a weight factor (Gardner, Mulvey, & Shaw, 1995). The weight factor is computed with the formula: 1/(value (deviance)/df (deviance)). In addition, given that Poisson and hierarchical regression does not account for censoring, we also included the observation period as a covariate.

The relation between program integrity and recidivism was investigated using Cox regression for the prevalence of recidivism and using Poisson and hierarchical regression for frequency and seriousness of recidivism. These analyses were only performed on the experimental group since program integrity of EQUIP could only be measured for that group.

## Results

### Effectiveness of EQUIP on recidivism

In Table 2, the percentages of recidivism at 6 months, 12 months, and 18 months are presented. When analyzing the prevalence of recidivism using Cox regression survival analysis, controlling for gender and seriousness of previous offense, we did not find a significant difference between the experimental and control group (OR = 1.65; CI 95% = .65 – 4.18; $p$ = .296). The covariate gender significantly predicted recidivism, with higher odds for boys to recidivate (OR = 3.11; $p$ = .031), but seriousness of previous offenses did not significantly predict recidivism (OR = 1.18; $p$ = .260). Figure 1 shows the prevalence of recidivism after release for the experimental and the control group.

Using Poisson regression, we did not find a significant difference between the experimental and control group in the frequency of recidivism ($B$ = –.931, $p$ = .090), though we did find a trend effect in favor of the control group. The control group showed a tendency to commit fewer offenses after release. Using hierarchical regression, no significant differences were found between the experimental and control group in the seriousness of recidivism ($B$ = .220, $p$ = .389). Covariates were not significantly related to frequency and seriousness of recidivism (all $p$ > .10), aside from the covariate gender that was significantly related to seriousness of recidivism ($B$ = .506, $p$ = .048), with higher odds for boys who commit more serious recidivism offenses.

### Program integrity of EQUIP

In the present study, the EQUIP program was implemented with low-to-moderate levels of composite program integrity ($M$ = 54%, SD = 7.6), ranging from 35% to 68%. We found that a quarter of the sample (23.6%) had an integrity score below 50%, while the majority of the sample (60.9%) had an integrity score between 50% and 60%. Only 15.5% of the

Table 2. Prevalence recidivism by observation period.

| | No period specified ($n = 133$) | | | | 6 months ($n = 132$) | | | | 12 months ($n = 126$) | | | | 18 months ($n = 83$) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | C | | E | | C | | E | | C | | E | | C | | E | |
| No recidivism | 18 | 78% | 62 | 56% | 21 | 91% | 93 | 85% | 19 | 86% | 71 | 68% | 11 | 82% | 50 | 71% |
| Recidivism | 5 | 22% | 48 | 44% | 2 | 9% | 16 | 15% | 3 | 14% | 33 | 32% | 2 | 18% | 20 | 29% |
| Total | 23 | 100% | 110 | 100% | 23 | 100% | 109 | 100% | 22 | 100% | 104 | 100% | 13 | 100% | 70 | 100% |

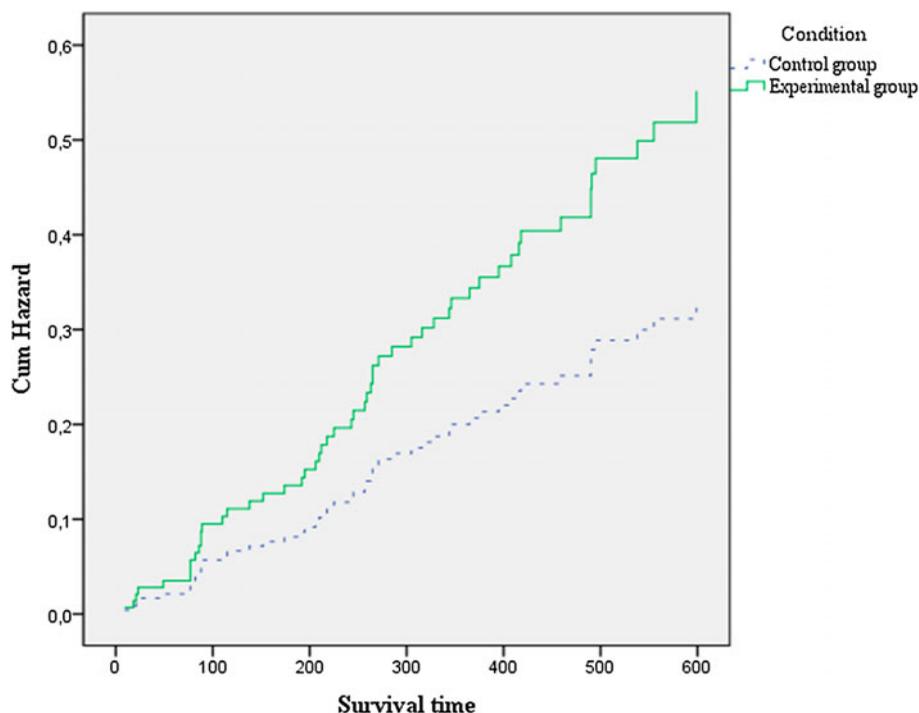C, control group; E, experimental group.

Figure 1. Prevalence of recidivism in the experimental and the control group.

sample had a program integrity score higher than 60%. When looking more specifically into the different aspects of integrity, we found facilities intended to implement half of the meetings of the program and one-third of the observed meetings were canceled. The meetings lasted 45 minutes instead of 60 minutes. Furthermore, adherence ranged on average from one-third to half of the content criteria being implemented during mutual help, anger management, social skills, and social decision-making meetings. Finally, the average participant responsiveness was about two-thirds of the highest possible score and little over half of the highest possible score for quality of delivery.

### Effects of program integrity of EQUIP on recidivism

For the experimental group, we examined whether program integrity strengthened the effectiveness of EQUIP on recidivism. We found that composite program integrity did not have a significant relation with the prevalence of recidivism (OR = 1.01; CI $95\% = .97$–1.04; $p = .796$), the frequency of recidivism ($B = .001$, $p = .962$), or seriousness of recidivism ($B = .018$, $p = .195$).

In addition to composite levels of program integrity, we also performed separate analyses on each of the program integrity facets and aspects, i.e., the facets exposure (frequency, cancelation, and duration), adherence to meetings (mutual help, anger management, social skills, social decision-making), participant responsiveness, and quality of delivery. The results of the different analyses all demonstrated a nonsignificant relationship between program integrity and recidivism (all $p > .10$).

## Discussion

The present study examined whether the cognitive behavioral intervention EQUIP for incarcerated adolescents would significantly reduce recidivism, and whether higher program integrity – the extent to which a program is implemented as intended – would strengthen the effectiveness of EQUIP on recidivism. The EQUIP program was implemented with low-to-moderate levels of program integrity in the Netherlands, and with these levels of program integrity, EQUIP was not effective in reducing recidivism. In addition, higher levels of program integrity, within the low-to-moderate range, did not strengthen the impact of the program on recidivism. High levels of integrity are a necessary pre-condition to draw valid conclusions regarding the effectiveness of intervention programs (Caroll et al., 2007; Mowbray et al., 2003); therefore, at present we are unable to draw final conclusions concerning the effectiveness of EQUIP.

In a related study we investigated the effect of EQUIP on cognitive distortions, social skills, moral judgment, and moral values (i.e., process outcomes; Helmond et al., 2012, 2014). In those studies, we demonstrated that both the EQUIP and the control group remained stable on cognitive distortions and moral judgment. However, while the EQUIP group remained stable in social skills and moral value evaluation, the control group showed a decrease in social skills and moral value evaluation. The present study revealed that the EQUIP program was not effective in reducing recidivism, as no differences were found between the experimental and control group on recidivism. Two possible explanations come to mind. The effects of EQUIP on these process outcomes were either too small to result in differences in recidivism or they were irrelevant as mediating variables to establish effects on recidivism. Brugman and Bink (2011) demonstrated that even though EQUIP helped to reduce cognitive distortions, this did not result in the expected reduction in recidivism. Together with the present study, these findings emphasize that improvements in process outcomes of cognitive behavioral programs cannot be expected to result in one-on-one reductions of recidivism.

When we compare the results of our study to other findings on the effectiveness of EQUIP on recidivism for juvenile offenders, we see that our results are comparable to findings by Brugman and Bink (2011). Both Dutch studies showed that EQUIP is not effective in terms of recidivism. On the contrary, both studies show a tendency for the control group to perform better in terms of recidivism outcomes than the experimental group. Brugman and Bink (2011) and Helmond et al. (2012, 2013, 2014) suggested that the lack of effectiveness on social cognitive outcomes and recidivism could be due to the fairly weak implementation of EQUIP. In these studies also the absence of a positive peer culture has been suggested as a potential explanation for the absence of the effectiveness of EQUIP. Neither Dutch studies, nor any other study that we know of, has replicated the initial promising effects of EQUIP for juvenile offenders on recidivism as demonstrated in the study by Leeman et al. (1993). These findings emphasize the importance of replicating initial promising findings, especially when the program is disseminated and no longer implemented by the program developers. In a previous study, we found that EQUIP was implemented with higher levels of integrity at the program developer site in comparison with nondeveloper sites (Helmond et al., 2013). This finding in combination with the Dutch results on the effectiveness of EQUIP raises the question whether EQUIP can be successfully disseminated at a large scale while maintaining the desired program outcomes. Currently, the EQUIP program is disseminated in the absence of a program integrity assurance system by the authors of the program. As a consequence, the

implementation quality of EQUIP cannot be monitored, leaving space for (non)deliberate personal adjustments in the implementation of the program. An example of such quality assurance system is given by the intervention Multisystemic Therapy that is facilitated by MST Services (2012a). The following quote gives an idea of the purpose of such a quality assurance system:

> MST is not a 'learn it and do it for the rest of your life' approach, the continuing support that MST services provides is crucial to the success of programs. Results are tracked and collectively shared with the greater MST professional community. Therapists working with these very challenging youths and families receive constant feedback, coaching and training (MST Services, 2012b).

The Ministry of Justice in the Netherlands has chosen for EQUIP as a general treatment program for all youth in juvenile correctional facilities in the Netherlands. Consequently, at time of our study all youth participated in the program irrespective of their level of risk to recidivate, while research shows that it is important to match treatment intensity to offenders' risk level to effectively reduce recidivism (Andrews & Dowden, 2006; Lowenkamp, Latessa, & Holsinger, 2006). The Risk-Need-Responsivity (RNR) model describes: (1) who to target (moderate and higher risk offenders), (2) what to target (criminogenic needs), and (3) how to target (apply certain general and specific strategies) (Andrews & Bonta, 2010). A wide body of research shows that programs adhering to the RNR model are more effective in reducing offender recidivism (Andrews & Bonta, 2010). Therefore, it is important that these RNR principles are used in correctional practice. There might even be a potential risk in not adhering to the RNR model. Some research indicates that treating low risk offenders with intensive treatment programs – intended for high risk offenders – might increase recidivism instead of reduce recidivism (Lowenkamp, Latessa, & Holsinger, 2006).

Another key finding was that EQUIP is not more effective in terms of lower prevalence, frequency, or seriousness of recidivism when implemented with higher – thus moderate instead of lower – levels of integrity. Do our findings implicate that program integrity is not important for the effectiveness of EQUIP? That conclusion cannot be drawn based on our findings given the relatively restricted range of program integrity in our study. Durlak and Dupre (2008) suggest that, as a rule of thumb, minimum levels of program integrity of 60% are needed to result in effective interventions. Interventions might be ineffective until a certain level of program integrity is achieved and may become effective only after surpassing that threshold level, suggesting that program integrity has a certain 'active range.' Because of the relatively restricted range, no information was available from participants who had received EQUIP with high levels of integrity. It could be that higher levels of program integrity need to be part of the sample to be able to establish a relationship between program integrity and effectiveness. An empirical example can be found in a study that used spline analysis on the relationship between childcare quality and child outcomes; the findings suggest there is no association between quality and outcomes at low quality levels, while there is a positive association between quality and outcomes at high quality levels (Burchinal, Xue, Tien, Auger, & Mashburn, 2011).

We need to know more about why some studies do show a relationship between program integrity and outcomes, while other studies do not. Several explanations come to mind: (1) some studies have evaluated programs that are ineffective in itself; therefore, it does not matter whether these programs are implemented with low or high levels of

integrity; (2) the program integrity source could influence the association between integrity and effectiveness, as self-evaluation can be positively biased and, consequently, unrealistically high levels of integrity are related to outcomes; (3) studies that include a limited number of program integrity facets might assess an unrepresentative program integrity score for the actual program implementation and therefore do not show the expected association between integrity and outcomes; and (4) studies include a restricted range of program integrity, meaning there is low variability to assess a relationship between integrity and outcomes.

We would like to make some recommendations for future studies examining the effectiveness of correctional treatment. It is important for treatment studies to assess both process and behavioral outcomes to get a better understanding of the working mechanisms of establishing behavioral change (Kazdin, 2007). This will provide insight whether treatment works in reducing criminal behavior and how the program helps to establish these effects, for instance by increasing, social skills and reducing cognitive distortions. Often correctional treatment studies study either process or behavioral outcomes.

Furthermore, it is important that treatment studies include a measure of program integrity because this can provide important understanding why programs are effective or ineffective. On the quality continuum of program integrity measures, objective measures, such as audio or video taping or direct observations, are rated the highest quality (Perepletchikova et al., 2007), but the great disadvantage of objective measures is that they are time-consuming. Videotaping without recognition could be a good option for correctional facilities, but we would like to point out that audio or video taping might be forbidden in a correctional setting or might negatively impact treatment process, because youth are not willing to expose themselves to taping. Even though objective integrity measures are time–consuming, we do recommend using them as they are more informative than self-evaluations that are often positively biased. It gives a more valid insight to how meetings are actually implemented.

There are a number of limitations of the present study that should be mentioned. A randomized design would have been preferable over a quasi-experimental design, as randomization of participants eliminates potential selection biases. However, randomized control trials are extremely difficult to accomplish within the juvenile justice system (Asscher, Dekovic, Van der Laan, Prins, & Van Arum, 2007). Consequently, relatively few randomized criminological intervention studies are conducted, especially in the Netherlands (Asscher et al., 2007; Farrington & Welsh, 2005; Wartna, 2009). Another concern is the small sample size of the study, more specifically of the control group. During our study, EQUIP was implemented as part of a nation-wide basic method called 'Youturn' for juvenile correctional facilities (Dienst Justitiële Inrichtingen, 2010). As a direct consequence of this policy, it was not possible to increase the size of our control group. All youth in Dutch juvenile correctional facilities were receiving the EQUIP program, leaving us without the possibility of creating a larger control group. The small sample size is also a consequence of the high levels of dropouts in our study. Dropouts were mainly the result of the referral process in the Dutch juvenile justice system and is part of the common situation in the Netherlands. Our attrition analysis demonstrated that youth with higher levels of cognitive distortions were more likely to remain part of the sample. Consequently, it is important to be careful in generalizing the results of our study to all youth in correctional facilities in the Netherlands because our sample represents only those youth who had more severe cognitive distortions. Despite these limitations, the present study has made an important contribution to the field with its elaborate program

integrity assessment by independent observers, the use of survival analyses, and by assessing the relation between program integrity and recidivism for a highly relevant clinical group of incarcerated juveniles in a real-life setting.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## References

Allen, J. D., Linnan, L. A., & Emmons, K. M. (2012). Fidelity and its relationship to implementation effectiveness, adaptation and dissemination. In R. C. Brownson, G. A. Colditz, & E. K. Proctor (Eds.), *Dissemination and implementation research in health* (pp. 281–304). New York, NY: Oxford University Press.

Andrews, D. A., & Bonta, J. (2010). Rehabilitating criminal justice policy and practice. *Psychology, Public Policy and Law*, *16*(1), 39–55. doi:10.1037/a0018362

Andrews, D. A., Bonta, J., & Wormith, J. S. (2006). The recent past and near future of risk and/or need assessment. *Crime and Delinquency*, *52*, 7–27. doi:10.1177/0011128705281756

Andrews, D. A., & Dowden, C. (2005). Managing correctional treatment for reduced recidivism: A meta-analytic review of programme integrity. *Legal and Criminological Psychology*, *10*, 173–187. doi:10.1348/135532505X36723

Andrews, D. A., & Dowden, C. (2006). Risk principle of case classification in correctional treatment: A meta-analytic investigation. *International Journal of Offender Therapy and Comparative Criminology*, *50*(1), 88–100. doi:10.1177/0306624X05282556

Asscher, J. J., Dekovic, 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*(2), 113–129. doi:10.1007/s11292-007-9028-y

Barnoski, R. (2004). *Outcome evaluation of Washington State's research-based programs for juvenile offenders*. Olympia: Washington State Institute for Public Policy.

Barriga, A. Q., Hawkins, M. A., & Camelia, C. R. T. (2008). Specificity of cognitive distortions to antisocial behaviours. *Criminal Behaviour and Mental Health*, *18*(2), 104–116. doi:10.1002/cbm.683

Beauchamp, M. H., & Anderson, V. (2010). SOCIAL: An integrative framework for the development of social skills. *Psychological Bulletin*, *136*(1), 39–64. doi:10.1037/a0017768

Brugman, D., & Bink, M. D. (2011). Effects of the EQUIP peer intervention program on self-serving cognitive distortions and recidivism among delinquent male adolescents. *Psychology, Crime & Law*, *17*, 345–358. doi:10.1080/10683160903257934

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

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*(1), 23–45. doi:10.1016/S0272-7358(97)00043-3

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*, 1–20.

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

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. doi:10.1007/s10464-008-9165-0

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*(1), 9–38. doi:10.1007/s11292-004-6460-0

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed poisson, and negative binomial models. *Psychological Bulletin*, *118*, 392–404. doi:10.1037/0033-2909.118.3.392

Gendreau, P., Goggin, C., & Smith, P. (1999). The forgotten issue in effective correctional treatment: Program implementation. *International Journal of Offender Therapy and Comparative Criminology*, *43*, 180–187. doi:10.1177/0306624X99432005

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.

Glick, B., & Gibbs, J. C. (2011). *Aggression replacement training: A comprehensive intervention for aggressive youth* (3rd ed.). Champaign, IL: Research Press.

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

Helmond, P. E., Overbeek, G. J., & 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. doi:10.1016/j.childyouth.2012.05.001

Helmond, P. E., Overbeek, G. J., & Brugman, D. (2013). A multiaspect program integrity assessment of the cognitive-behavioral program EQUIP for incarcerated offenders. *International Journal of Offender Therapy and Comparative Criminology*, *58*(10), 1186–1204. doi:10.1177/0306624X13494171

Helmond, P. E., Overbeek, G. J., & 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. doi:10.1016/j.childyouth.2014.01.022

Helmond, P. E., Overbeek, G. J., Brugman, D., & Gibbs, J. C. (2014). A meta-analysis on cognitive distortions and externalizing problem behavior: Associations, moderators, and treatment effectiveness. *Criminal Justice and Behavior*, Advance online publication. doi:10.1177/0093854814552842

Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York, NY: Routledge.

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology 3*, 1–27.

Kleinbaum, D. G., & Klein, M. (2005). *Survival analysis: A self-learning text (statistics for biology and health)*. New York: Springer.

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.

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. doi:10.1007/s11292-005-3541-7

Leeman, L.W., Gibbs, J. C., & Fuller, D. (1993). Evaluation of a multi-component group treatment program for juvenile delinquents. *Aggressive Behavior*, *19*, 281–292. doi:10.1002/1098-2337(1993)19:4<281::AID-AB2480190404>3.0.CO;2-W

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. doi:10.1002/jcop.20018

Lillehoj, C.J., 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. doi:10.1177/1090198103260514

Lipsey, M. W. (2009). The primary factors that characterize effective interventions with juvenile offenders: A meta-analytic overview. *Victims & Offenders*, *4*(2), 124–147. doi:10.1080/15564880802612573

Lösel, F., & Beelmann, A. (2003). Effects of child skills training in preventing antisocial behavior: A systematic review of randomized evaluations. *The Annals of the American Academy of Political and Social Science*, *587*, 84–109.

Lowenkamp, C. T., Latessa, E. J., & Holsinger, A. M. (2006). The risk principle in action: What have we learned from 13,676 offenders and 97 correctional programs? *Crime & Delinquency*, *52* (*1*), 77–93. doi:10.1177/0011128705281747

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. doi:10.1111/j.1745-9133.2006.00388.x

Lowenkamp, C. T., Makarios, M. D., Latessa, E. J., Lemke, R., & Smith, P. (2010). Community corrections facilities for juvenile offenders in Ohio: An examination of treatment integrity and recidivism. *Criminal Justice and Behavior*, *37*, 695–708. doi:10.1177/0093854810363721

McGuire, J. (2000). Development of a program logic model to assist evaluation. In L. L. Motiuk & R. C. Serin (Eds.), *Compendium 2000 on effective correctional programming* (pp. 208–220). Ottawa: Correctional Services of Canada.

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. doi:10.1177/109821400302400303

MST Services. (2012a, November, 4). *Overview: The Multisystemic Therapy (MST) QA/QI program*. Retrieved from htttp://www.mstinstitute.org/qa_program

MST Services. (2012b, November, 4). *MST services*. Retrieved from http://mstservices.com/indexphp/mst-services

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. doi:10.1080/10683160500255703

Nas, C. N., Brugman, D., & Koops, W. (2008). Measuring self-serving cognitive distortions with the "How I Think" Questionnaire. *European Journal of Psychological Assessment*, *24*(3), 181–189. doi:10.1027/1015-5759.24.3.181

Perepletchikova, F. (2011). On the topic of treatment integrity. *Clinical Psychology: Science and Practice*, *18*, 148–153. doi:10.1111/j.1468-2850.2011.01246.x

Perepletchikova, F., Treat, T. A., & Kazdin, A. E. (2007). Treatment integrity in psychotherapy research: Analysis of the studies and examination of the associated factors. *Journal of Consulting and Clinical Psychology*, *75*, 829–841. doi:10.1037/0022-006X.75.6.829

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

Raaijmakers, A. W., Engels, R. C. M. E., & Van Hoof, A. (2005). Delinquency and moral reasoning in adolescence and young adulthood. *International Journal of Behavioral Development*, *29*, 247–258. doi:10.1177/01650250544000035

Rasbash, J., Charlton, C., Browne, W. J., Healy, M., & Cameron, B. (2010). MLwiN (Version 2.21). Bristol: Centre for Multilevel Modelling, University of Bristol.

Stams, G. J., Brugman, D., Dekovic, M., Van Rosmalen, L., Van der Laan, P., & Gibbs, J. C. (2006). The moral judgment of juvenile delinquents: A meta-analysis. *Journal of Abnormal Child Psychology*, *34*, 692–708. doi:10.1007/s10802-006-9056-5

Sullivan, C. J., & Latessa, E. (2011). The coproduction of outcomes: An integrated assessment of youth and program effects on recidivism. *Youth Violence and Juvenile Justice*, *9*, 191–206. doi:10.1177/1541204010393754

Van Vugt, E., Gibbs, J., Stams, G. J., Bijleveld, C., Hendriks, J., & Van der Laan, P. (2011). Moral development and recidivism: A meta-analysis. *International Journal of Offender Therapy and Comparative Criminology*, *55*, 1234–1250. doi:10.1177/0306624X11396441

Vartuli, S., & Rohs, J. (2009). Assurance of outcome evaluation: Curriculum fidelity. *Journal of Research in Childhood Education*, *23*, 502–512. doi:10.1080/02568540909594677

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

Wartna, B. S. J. (2009). *In de oude fout* [English: The same mistake]. The Hague*:* Research and Documentation Centre.

Wartna, B. S. J., Blom, M., & Tollenaar, N. (2011). *The Dutch recidivism monitor*. The Hague: Research and Documentation Centre.

Wartna, B. S. J., El Harbachi, S., & Van der Laan, A. M. (2005). *Jong vast. Een cijfermatig overzicht van de strafrechtelijke recidive van ex-pupillen van justitiële jeugdinrichtingen* [English: Young and incarcerated. A statistical overview of the criminal recidivism of former inmates of juvenile detention]. The Hague*:* Research and Documentation Centre.