Effects of Awareness Programs on Juvenile Delinquency: A Three-Level Meta-Analysis

van der Put, C.E.; Boekhout van Solinge, N.F.; Stams, G.J.; Hoeve, M.; Assink, M.

DOI
10.1177/0306624X20909239

Publication date
2021

Document Version
Author accepted manuscript

Published in
International Journal of Offender Therapy and Comparative Criminology

Effects of Awareness Programs on Juvenile Delinquency: A Three-Level Meta-Analysis

Claudia E. van der Put, Noelle F. Boekhout van Solinge, Geert Jan J. M. Stams, Machteld Hoeve, & Mark Assink

Journal: International Journal of Offender Therapy and Comparative Criminology
Accepted: January 31, 2020

Effects of awareness programs on juvenile delinquency: A three-level meta-analysis

Claudia E. van der Put[a], Noelle F. Boekhout van Solinge[a], Geert Jan J. M. Stams[a], Machteld Hoeve[a], & Mark Assink[a]

University of Amsterdam

Author Note

[a] Research Institute of Child Development and Education, University of Amsterdam, Nieuwe Achtergracht 127, 1018 WS, Amsterdam, the Netherlands.

Declarations of interest: none.

Correspondence concerning this article should be addressed to Claudia E. van der Put, Research Institute of Child Development and Education, University of Amsterdam, P.O. Box 15780, 1001 NG, Amsterdam, The Netherlands. Phone: 0031-(0)6-26086910. E-mail: C.E.vanderPut@UvA.nl

Abstract

Juvenile awareness programs, such as Scared Straight, remain in use despite the finding that these programs provoke rather than prevent delinquency. The aim of this study was to examine what program components are associated with program effectiveness, which is important for improving these programs. A three-level meta-analysis was conducted. A literature search yielded 13 independent studies (N=1,536) from which 88 effect sizes could be extracted. A non-significant overall effect was found (d=0.10), indicating that juvenile awareness programs have no effect on offending behavior and other outcomes that are related to delinquency. No significant moderator effects were found for program components. The moderator analyses revealed that juvenile awareness programs are effective in reducing antisocial attitudes (d=0.46), which has not been meta-analytically studied before. Furthermore, larger effects were found as follow-up length increased. These results show a more nuanced view on the effectiveness of juvenile awareness programs is necessary.

Keywords: crime prevention; juvenile awareness programs; effectiveness; effective components; meta-analysis; Scared Straight

Effects of awareness programs on juvenile delinquency: a three-level meta-analysis

In juvenile awareness programs, juvenile delinquents or youths at-risk for delinquency participate in an organized visit to a prison facility (Petrosino, Turpin-Petrosino, Hollis-Peel, & Lavenberg, 2013). The main goal of such programs is to deter youth from future criminal behavior by exposing them to realistic depictions of life in prison and presentations held by inmates. Petrosino et al., (2013) concluded in their meta-analysis that juvenile awareness programs provoke rather than prevent delinquency. Nevertheless, these programs remain in use worldwide. In the Netherlands, juvenile awareness programs have even increased in popularity, though in a slightly different form in which ex-prisoners visit schools rather than children visiting prisons (Van Kempen, Van Rooij, & Rovers, 2010). In these adapted programs, which have a similar crime prevention goal to the original programs, ex-prisoners share their life stories and describe the choices they made that led to imprisonment.

In this study, we aimed to gain knowledge on what program components are (positively or negatively) associated with program effectiveness, and what components are not related to effectiveness. This knowledge is important for improving these programs by adding effective components to existing programs and/or eliminating ineffective components from those programs. For this purpose, the meta-analysis of Petrosino and colleagues (2013) was updated and extended by (a) examining how individual program components and study characteristics influence program effectiveness, (b) including not only studies on prison tour programs, but also studies on programs in which (ex-)prisoners visit schools, (c) using a three-level meta-analytic design allowing the extraction of multiple effect sizes from individual primary studies, and (d) examining program effects on not only delinquent behavior, but also pro-delinquent attitudes, attitudes toward punishment, and risk factors for delinquency.

Juvenile awareness programs were first developed around 1970 in the United States, with “Scared-Straight” being the most well-known. In the Scared-Straight program, at-risk youths and

juvenile delinquents were taken to a prison facility where inmates serving life sentences held an aggressive presentation emphasizing the harsh realities of a life in prison. These presentations were mainly rude and confrontational, and included exaggerated stories about rape and murder (Finckenauer, 1982). The program became highly popular and was soon picked up by multiple states as well as other countries, such as Australia, Canada, United Kingdom, and Norway (Petrosino et al., 2013).

The popularity of the Scared-Straight program can be explained by several factors. First, in 1979, an Oscar-winning documentary about the program was aired, claiming high success rates in preventing or deterring youths from delinquency (Cook & Spirrison, 1992; Finckenauer, 1982). Second, the program fits with some commonly held notions about how to prevent crime, and was seen as easy to implement at low costs (Petrosino et al., 2013). Third, the program provides an opportunity for inmates to contribute to society in a positive way by preventing youth to end up like themselves (Finckenauer, 1982). Besides the intended goal of preventing juveniles from committing offenses, there are possible benefits for (ex-)prisoners as well. Research showed that helping others can facilitate the recovery and reintegration of (ex-)prisoners (Lebel, 2007).

The supposed underlying mechanism of juvenile awareness programs is grounded in deterrence theory. Deterrence is based on the theoretical notion that offenders learn from the negative experience of (fear of) punishment, and thus will deter from further offensive behavior. This criminological theory partly shaped the criminal system of the United States and various other countries (Pratt, Cullen, Blevins, Daigle, & Madensen, 2006). Both the certainty and severity of punishment are perceived to be important concepts in preventing and deterring people from involvement in crime (Paternoster, 2010). Following this theory, the rationale of juvenile awareness programs is that experiences and realistic depictions of life in prison scare at risk youths and juvenile offenders deterring them from future involvement in crime. However, the claimed positive results of juvenile awareness programs, such as Scared-Straight, (i.e., Muhammed, 1999) were quickly rejected

in multiple experimental studies (GERP&DC, 1979; Lewis, 1983; Yarborough, 1979). In fact, some studies even concluded that the program produced an adverse effect, because delinquency rates were significantly higher among program participants than among non-participants (Finckenauer, 1982; Michigan Department of Corrections, 1967).

In 2002, Petrosino, Turpin-Petrosino, and Buehler (2002) performed the first review on the effectiveness of Scared-Straight and similar programs. The authors synthesized the results of nine randomized trials in which delinquency rates of participants were compared to delinquency rates of non-participants. The results showed that programs such as Scared Straight generally increase offending levels with a rate of 1% to 28% in the experimental group relative to a no-treatment control group. Petrosino et al. concluded that Scared-Straight and similar programs result in harmful effects for participants, as they found higher delinquency rates in youth exposed to these programs than in non-exposed youth. Despite these negative results, juvenile awareness programs remained popular. In 2011, a weekly television show entitled 'Beyond Scared Straight' led to renewed interest in Scared-Straight-type programs. In 2013, Petrosino and colleagues updated their earlier meta-analysis that was published in 2002, but in the search procedure, no new trials were found that were eligible for inclusion (Petrosino, Turpin-Petrosino, Hollis-Peel, & Lavenberg, 2013). In contrast to the 2002 review in which no pooled or overall effect was calculated, the pooled effect in the 2013 review showed that officially measured criminal behavior was more prevalent in the treatment group than in the control group (OR = 1.68) indicating a negative overall program effect. The authors concluded once again that juvenile awareness programs increase instead of decrease delinquency (Petrosino et al., 2013).

The present review was directed at updating and extending the meta-analysis of Petrosino and colleagues (2013) for several reasons. First, the main objective was to enhance knowledge on how individual components of juvenile awareness programs are related to the effectiveness of these programs. Some awareness programs are for instance highly confrontational, whereas other

programs are more educational in nature (Finckenauer, Gravin, Hovland, & Storvoll, 1999; Lundman, 1993). It is not yet clear how the components of education and confrontation are associated with program effectiveness. Therefore, we tested different program components as moderators of the (mean or overall) effect of juvenile awareness programs. Second, we aimed to synthesize studies on the effect of prison tour programs as well as studies on the effect of programs in which (ex-)prisoners visit schools. Third, Petrosino and colleagues (2013) used a classic meta-analytic design in which one effect size per primary study was included. Specifically, they only examined immediate post-treatment effects and disregarded any follow-up effects that were reported in primary studies. In the present study, a three-level meta-analytic design was used, in which effect size dependency is accounted for, so that multiple effect sizes can be extracted from individual primary studies. In this way, all relevant information can be preserved and maximum statistical power in the analyses can be achieved (Assink & Wibbelink, 2016). By applying this three-level approach to meta-analysis, all follow-up effects reported in the primary studies could be included and synthesized. Fourth, Petrosino and colleagues only included effect sizes based on officially recorded offending, whereas we included effect sizes based on either officially-recorded delinquency or self-reported delinquency. Both forms of delinquency measurements have their strengths and limitations, but specifically using official records involves the risk of underestimating the actual number of criminal acts, as there is more criminality than is registered in the official systems. Therefore, it is relevant to also include studies based on self-reported delinquency. Finally, we included and synthesized both delinquency outcomes and other outcomes associated with delinquency, such as attitudes toward punishment, pro-delinquent attitudes and other risk factors for delinquency,

As for the individual program components that may be associated with program effectiveness, we examined the effect of the following components: rap sessions (confrontational sessions in which adult inmates share graphic stories about prison life with juveniles), confining

juveniles (locking up of juveniles in a cell for a short time), providing juveniles a guided tour of a prison facility, visitation (juveniles are (body-)searched and personal items are taken away and held during the program), dialogue (discussion with inmates), counseling, and attending presentations, for example about drug and alcohol use or other topics related to delinquency. Regarding program characteristics, we examined: the program duration (in hours), the level of confrontation (high [e.g., verbal intimidation, confrontational lectures, strutting, yelling, or bullying], moderate [some provocation and baiting of the juveniles by inmates, but not as much as in confrontational programs], or low [educational instead of confrontational sessions, in which (ex-)prisoners shared their life stories and describe the choices they made that led to imprisonment], whether a program is offered inside or outside a prison facility, and whether or not parents are involved in a program. We also examined the effect of the following study design characteristics: sample size, dropout percentage, sample type (delinquents, non-delinquents), sex of the sample (boys, girls, both), mean age of the sample, percentage of racial minorities in the sample, research design of the study (randomized controlled trial versus quasi-experimental), follow-up length of an assessment (in months), type of measurement (official records, self-report, parent-report), and type of outcome (level of delinquency, attitudes, other non-attitude risk factors for delinquency).

Method

Sample of Studies

For selecting relevant studies, several criteria were formulated. First, we selected studies that examined the effects of (a) programs involving organized visits to prison facilities for juvenile delinquents or youths at-risk of becoming delinquent with the aim to prevent or deter them from juvenile delinquency, and (b) programs in which juveniles come into contact with prisoners or ex-prisoners visiting schools, with the aim to increase (prison) awareness and thereby prevent delinquency. Second, samples of primary studies had to consist of juvenile delinquents (aged about 12 to 20 years), or youths at risk for delinquency. The latter group was defined as troubled youths

who had not (yet) been officially adjudicated as delinquents (Petrosino et al, 2013), or as youths with an increased risk of delinquency due to the presence of one or more risk factors, such as a disadvantaged neighborhood, a low family social economic status, and a low educational level. Third, primary studies had to report on levels of offending behavior after a program was ended (assessed using official records and/or self-report instruments), attitudes toward delinquency and/or punishment, or levels of risk factors for delinquency (e.g., school dropout, having delinquent friends, impulsiveness). Fourth, the study had to be experimental (i.e., a treatment group was compared to a comparison group of juveniles who did not participate in a juvenile awareness program). Fifth, randomized controlled trials (RCTs) and high quality quasi-experimental studies were included. A quasi-experimental study was considered 'high-quality' in case (a) participants were assigned to conditions by odd/even assignment and/or (b) an adequate matching procedure was used and group equality was measured at pretest. Although RCTs can be regarded as the "golden standard" in effectiveness studies (Farrington, 2003),we decided to include both RCTs and quasi-experimental studies, as juvenile awareness programs have been examined in RCTs to a limited extent, and because of their clinical representativeness (Shadish, Navarro, Matt, & Philips, 2000). Besides, by including both RCTs and quasi-experimental studies, the power in the statistical analyses can be increased, generalizability of the results is enhanced, and relevant results may be obtained that could have been missed in a synthesis of merely RCT's. Sixth, both published and unpublished studies, such as government reports and doctoral dissertations, were included to reduce the risk of (publication) bias in the results and to increase the representativeness and generalizability of our study findings. Seventh, studies had to report actual effect sizes, or sufficient statistical information that is required for calculating an effect size manually (e.g., contingency tables, mean scores and standard deviations, or proportions). Finally, no date restrictions were set in our search, and articles had to be written in English.

Search Strategy

Until October 2017, we searched for articles, book chapters, dissertations, and reports in the following five electronic databases: PubMed, PsycINFO, Google Scholar, Google, and Web of Science. In searching these databases, the following keywords were used in different combinations reflecting the study design, type of program, study outcomes, and participants we were interested in: 'experiment', 'quasi-experiment', 'randomized control*', 'RCT', 'evaluat*', 'effect*', 'prison awareness', 'prison aversion', 'juvenile awareness', 'delinquen*', 'criminal*', 'attitude', 'prisoner*', and 'detainee*'. To assess the retrieved studies against all formulated inclusion criteria, we read titles, abstracts and, if necessary, full article texts.

Next, studies were searched by screening the reference list of the meta-analysis of Petrosino and colleagues (2013). We also contacted several scholars by email to request for published and unpublished studies that would meet our inclusion criteria. Finally, we screened the reference list of each primary study that was eligible for inclusion for additional relevant studies. The full search procedure is depicted in the flow diagram presented in Figure 1. The search yielded 28 relevant studies of which 13 studies met the inclusion criteria.

Study Coding

A coding scheme was developed for coding all study design, sample, outcome, and intervention characteristics that we were interested in. In developing this scheme, the guidelines of Lipsey and Wilson (2001) were followed. In the first coding round, the first and second author independently coded seven randomly selected studies that were eligible for inclusion. All codings of both raters were compared, and coding inconsistencies were discussed and resolved until both authors agreed on all final coding decisions. In case consensus on a coding could not be reached, the last author was consulted and acted as an arbitrator. This first round ended with improving the coding procedure and code form, where appropriately. In the second coding round, the second author coded the remaining six studies.

Regarding the *study design characteristics*, we coded publication year, publication type (peer-reviewed article, research report, dissertation, book, government report), and research design (randomized controlled trial, quasi-experimental study). Next, *sample characteristics* were coded. Information was collected on the total sample size, dropout percentage, sample type (non-delinquents, delinquents), sex of participants in the sample (boys, girls, both boys and girls), the mean age of participants in the sample, age range (oldest – youngest participant), and percentage of racial minorities in the sample. As for the *outcome characteristics*, we coded follow-up length of an assessment (in months), type of measurement (official records, self-report, parent-report), and type of outcome measure (delinquency [frequency or severity], attitude toward crime [for example attitude toward obeying the law], attitude toward punishment [for example attitude toward prison and the practice of punishing criminals], and other non-attitude risk factors for delinquency [for example, school dropout, having delinquent friends, impulsiveness, hostility, low inhibition, and poor family relationship].

Regarding *intervention characteristics*, we coded the duration of the program (in hours), the level of confrontation used in the program (high [e.g., verbal intimidation, confrontational lectures, strutting, yelling, or bullying], moderate [some provocation and baiting of the juveniles by inmates, but not as much as in confrontational programs], or low [educational instead of confrontational sessions, in which (ex-)prisoners shared their life stories and describe the choices they made that led to imprisonment]), whether the program was offered inside or outside a prison facility, whether parents were involved in the program (yes/no), and the intervention form (only group sessions or both individual and group sessions). As for *intervention content*, we coded whether the following components were present or absent in a juvenile awareness program: rap session in which adult inmates shared graphic stories about prison life with juveniles, temporarily confinement of juveniles, providing a guided tour of a prison facility, visitation (juveniles are (body-)searched and personal items are taken away and held during the program), dialogue (discussion with inmates), counseling,

and attending a presentation, for example about drug and alcohol use or other topics related to delinquency.

Calculation of Effect Sizes

All effects of a juvenile awareness program as reported in each primary study were transformed into Cohen's *d*, as participants attending a program in the experimental group were compared to non-attending participants in the control group. For calculating Cohen's *d* based on the information that was reported in primary studies, formulas of Ferguson (1966), Lipsey and Wilson (2001), and Rosenthal (1994) were used. In most cases, cohen's *d* was calculated by using proportions, or means and standard deviations. When raw outcome data were not reported in a study, we transformed a test-statistic (*F* value, *z* value, χ^2 value or *t* value) into cohen's *d*.

In extracting and calculating effect sizes, it was important that program effects were properly expressed in *d* values, which could be either positive or negative. A positive *d*-value indicated lower levels of delinquency, less negative attitudes, or lower levels of risk factors in the experimental group relative to the control group. On the other hand, a negative *d* value indicated higher levels of delinquency, more negative attitudes, or higher levels of risk factors in the experimental group than in the control group.

All data were entered in SPSS version 24. In preparing the dataset for the analyses, the continuous variables that were to be tested as moderators were centered around their mean, and each category of potential moderating discrete variables were recoded into a dummy variable. Next, standardized scores were calculated to search for outliers, since extreme values of effect sizes may have a disproportionate influence on the results of the statistical analyses. No *d* values were found with a *Z* score above 3.29 or lower than -3.29 (Tabachnik & Fidell, 2013), so no outliers were identified.

Statistical Analyses

A three-level meta-analysis was conducted for estimating the overall effect of juvenile awareness programs, and to test variables as moderators of this effect. In this approach, there is no need for selecting or aggregating outcomes reported in primary studies, as dependency between outcomes is modeled in this three-level approach to meta-analysis. This implies that all relevant effect sizes can be extracted from each primary study (see, for instance, Assink & Wibbelink, 2016). In contrast, Petrosino et al. (2013) used a classical two-level approach, implying that only one effect size could be extracted from an included study. After all, a key assumption in meta-analysis is that effect sizes need to be independent, so extracting multiple effect sizes based on the same sample would violate this assumption in traditional approaches to meta-analysis (see, for instance, Lipsey & Wilson, 2001).

To deal with dependency of effect sizes in meta-analysis, a multilevel random effects model can be used for calculating a combined effect size and conducting moderator analyses (Hox, 2002; Van den Noortgate & Onghena, 2003). As noted by Van den Noortgate and Onghena (2003), who compared the multilevel approach to the traditional fixed-effects approaches, the former seems to be superior in conducting meta-analyses. They also concluded that for models without moderators, the results of the multilevel approach did not differ substantially from results of the traditional random-effects approaches.

In a three-level meta-analytic model, three sources of variance are modeled: random sampling variation of observed effect sizes (level 1), variance between effect sizes within studies (level 2), and variance between studies (level 3) (Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2013; 2015). For estimating the overall effect, we built an intercept-only model without any covariates. In this model, the intercept represented the overall effect. Next, we examined the significance of the variance distributed at levels 2 and 3 of the three-level model by performing two separate one-tailed log-likelihood-ratio tests. If the variance distributed at one of these levels is significant, the distribution of effect sizes can assumed to be heterogeneous, implying that moderator

analyses can be conducted to examine what study design or intervention characteristics may explain level 2 and/or level 3 variance. In performing these moderator analyses, the intercept-only model was extended by adding covariates (i.e., the potential moderating variables that we coded) to the model. In determining the significance of all estimated regression coefficients, the Knapp and Hartung adjustment (2003) was applied, implying that the *t* and *F* distribution (rather than the *z* distribution) were used. This adjustment reduces the number of unjustified significant results, as this is a problem with using the *z* distribution (see, for instance, Li, Shi, & Roth, 1994; Ziegler, Koch, & Victor, 2001).

All analyses were performed using the statistical software package R (version 3.4.2) and the metafor-package (Viechtbauer, 2010). The manual written by Assink and Wibbelink (2016) was followed in writing the syntaxes. Results were considered 'significant' when the 5% significance level was reached, and 'non-significant trend' when the 10% significance level was reached.

Publication Bias

One of the problems in meta-analysis is publication bias, also referred to as the 'file drawer problem' by Rosenthal (1995), which implies that studies producing non-significant or negative results are less likely to be published than studies producing positive and significant results. Therefore, the studies included in a review may not be an adequate representation of all available studies relevant to a particular subject, and thus the results may be biased. To examine whether the results of the present meta-analysis were affected by (different forms of) bias, we conducted the non-parametric and funnel-plot based trim-and-fill analysis as described by Duval and Tweedie (2000a, 2000b). In this analysis, the symmetry of the funnel plot – in which effect sizes are plotted against their standard error – is tested. Bias may be present if the funnel is asymmetric. In case of an asymmetric funnel, the symmetry can be restored by imputing "missing" effect sizes that are estimated on the basis of existing effect sizes in the dataset. After imputing the "missing" effect sizes, an adjusted overall effect can be estimated.

Results

Descriptive Characteristics

In the present meta-analysis, $k = 13$ studies published between 1967 and 1992 (median publication year is 1981) were included. In total, 88 effect sizes could be extracted from these studies. The included studies comprised peer-reviewed articles ($k = 5$), research reports ($k = 3$), dissertations ($k = 3$), one book ($k = 1$), and one unpublished government report ($k = 1$). The total number of participants examined in all included studies was $N = 1{,}536$ youths and young adults, and the sample size of primary studies ranged from 28 to 300. The age of participants ranged from 7 to 20 years, and the mean age of participants was 15.48 years ($SD = 0.74$). The mean percentage of boys in primary study samples was 92.1% ($SD = 6.51$). All included studies were conducted in the United States. Appendix A presents an overview of all included studies with several study characteristics.

Overall Effect and Heterogeneity in Effect Sizes

A non-significant overall effect was found of $d = 0.099$; 95% CI [-0.040, 0.238], $t(87) = 1.420$, $p = 0.159$ (see Table 2). The two log-likelihood-ratio tests revealed significant variance at both level 2 ($\chi^2(1) = 13.628$, $p < .001$; one-sided) and level 3 ($\chi^2(1) = 7.894$, $p < .010$; one-sided) of the three-level meta-analytic model. Of the total variance, 31.0% and 29.0% were distributed at levels 2 and 3 respectively, and 40.0% was the percentage of sampling variance that was calculated using the formula of Cheung (2014, p. 15). The results of the trim-and-fill analysis showed that bias may be present in the data, because the distribution of effect sizes was asymmetrical. Figure 2 shows that 4 effect sizes (from 3 studies) had to be imputed in the left side of the funnel to restore the symmetry of the effect size distribution. After imputing these 4 effect sizes and re-estimating the overall effect, a lower effect was obtained of $d = 0.020$; 95% CI [-0.146, 0.186], $t(91) = 0.241$, $p = 0.810$, $\Delta\, d = 0.079$.

Moderator Effects

Table 2 shows the results of the moderator analyses. As for the study characteristics, a significant moderating effect was found for publication year. The overall effect increased as studies were published more recently (β_1 = 0.038, p < .05). Further, a significant moderating effect was found for type of measurement. Official record assessments produced a lower mean effect (d = -0.055) than self-report assessments (d = 0.320). Further, the type of outcome significantly moderated the overall effect. The mean effect was larger when delinquency risk factors (d = 0.197), attitudes toward delinquency (d = 0.460), and attitudes toward punishment (d = 0.347) were assessed relative to the assessment of delinquency (d = -0.019). Further, the overall effect significantly increased as the follow-up duration increased (β_1 = 0.042, p <.05). To test this moderator separately for delinquency outcomes, delinquency attitudes and risk factors for delinquency, three additional moderator analyses were performed. In these moderator analyses, the follow-up length was tested as a binary categorical moderator with a category for a follow-up length of 0 to 6 months and a category for a follow-up length of more than 6 months. The results are shown in Table 3. All moderator effects yielded significant results (p < .05), meaning that for all outcome types larger effects were found for a longer follow-up length. The mean effects for assessments with a follow-up length of more than 6 months were positive, although this mean effect did not significantly deviate from zero for delinquency outcomes (see Table 3). Finally, no moderating effects were found for intervention characteristics. The presence (or absence) of specific program components, the level of confrontation, and the duration and setting of programs did not significantly explain variance in effect sizes.

Discussion

Juvenile awareness programs remain in use despite the finding that these programs provoke rather than prevent delinquency (Petrosino et al., 1002, 2013). The aim of the present study was to contribute to a better understanding of why these programs are not effective in preventing juvenile delinquency by unravelling program components that are negatively associated with program

effectiveness from components that are positively associated with program effectiveness. A literature search yielded 13 independent studies (N =1,536) from which 88 effect sizes could be extracted. A three-level meta-analysis was conducted to determine the overall effect of juvenile awareness programs on delinquent behavior, pro-delinquent attitudes, attitudes toward punishment, and other risk factors for delinquency (e.g., school dropout, having delinquent friends, being hostile). A non-significant overall effect was found (d = 0.099), indicating that juvenile awareness programs have no positive or negative effect on offending behavior and other outcomes that are related to delinquency. The results of the trim-and-fill-analysis suggested that bias may have been present in the data, and therefore a "corrected" overall effect was estimated, resulting in a smaller and also non-significant effect size of d = 0.020. Because there are several methodological shortcomings of the trim-and-fill method (see the Limitations section), this corrected effect size should not be interpreted as an estimate of the true effect size, but only as an indicator of (possible) bias in the data.

Program effects were significantly larger when delinquency risk factors (d = 0.197), attitudes toward delinquency (d = 0.460), and attitudes toward punishment (d = 0.347) were measured, relative to direct measures of delinquency (d = -0.019). So, juvenile awareness programs have a significant positive effect on attitudes toward delinquency and attitudes toward punishment. There may also be a positive effect of juvenile awareness programs on levels of risk factors for delinquency, such as school dropout, having delinquent friends, and being hostile, as we found a non-significant trend. The larger effects of programs on attitudes may be explained by the observation that a change in attitude generally precedes a change in behavior (see for example the behavior change model of De Vries, 2017). Consequently, a follow-up of substantial length may be required to measure a change in delinquent behavior, whereas a follow-up of rather short length may be sufficient to measure a change in attitude. In line with this reasoning, we found that juvenile awareness programs are more effective as the follow-up length increased, indicating that relatively long follow-up periods may be necessary for properly measuring program effects on delinquent

behavior. In fact, the mean follow-up duration in the included studies was only 5.12 months (SD = 4.59).

There are two main differences between the meta-analysis of Petrosino and colleagues (2013) and the present study. First, Petrosino and colleagues only synthesized immediate post-treatment effects (that is, 'first-effects') even though primary studies reported multiple follow-up measurements. In contrast, all post-treatment effects were included and synthesized in the present meta-analysis. Second, Petrosino and colleagues included only effect sizes that were based on officially recorded offending behavior, whereas the present study also included effect sizes that were based on self-reported offending behavior as well as other outcomes that are related to delinquency, such as risk factors for delinquency, attitudes toward punishment, and pro-delinquent attitudes. Because we used a three-level meta-analytic design, all effect sizes could be extracted from each included primary study, implying that the maximum relevant information could be retained. Consequently, we believe that the finding of Petrosino and colleagues, that juvenile awareness programs increase instead of prevent delinquency, deserves to be nuanced for two reasons. For one, this is the first meta-analysis showing that juvenile awareness programs are effective in reducing antisocial attitudes. Perhaps these programs even reduce levels of risk factors for delinquency, but as a non-significant trend was found, this needs to be examined in future research before firm conclusions can be drawn. Second, program effects seem to increase as follow-up length increases. This was found in the analysis of both delinquency and antisocial attitude outcomes, with a mean positive effect for assessments that were performed with a follow-up length of more than 6 months. However, for delinquency outcomes, this mean effect did not significantly deviate from zero.

Regarding the effectiveness of program components, such as level of confrontation, involvement of parents, and individual counseling, no significant moderating effects were found, suggesting that these components are about equally associated with program effectiveness. So, for example, there is no evidence that juvenile awareness programs with high levels of confrontation

(e.g., telling stories about rape and murder to depict the harsh reality of prison life) perform better or worse than programs with low levels of confrontation, which may be more educational in nature. Based on the current results, we have no better understanding of why juvenile awareness programs are not effective and how these programs can be improved, because no individual components were significantly associated with the effectiveness of juvenile awareness programs. However, the current results do show a somewhat more positive picture of program effectiveness than the previous meta-analysis (Petrosino et al, 2013), given the improvements in antisocial attitudes, and the positive moderating effect of follow-up length.

Several limitations need to be addressed. First, only a small amount of empirical studies on the effect of juvenile awareness programs on offending behavior and/or other outcomes related to delinquency is available. Therefore, only 13 studies were included in the current review. On the other hand, the three-level approach to meta-analysis allowed us to extract multiple effect sizes from most included studies. An important advantage of the three-level approach is that all relevant information produced in primary studies can be preserved and maximum statistical power can be achieved (Assink & Wibbelink, 2016). Second, we were not able to find any recent published or unpublished study on the effectiveness of a juvenile awareness program. The included studies were published between 1967 and 1992, indicating that many years have passed since the included studies were published and that the time frame in which these studies were performed was rather long. Most probably, this affects the generalizability of the current results, because over time, social and societal changes took place that have influenced the prevalence of crime and the attitude toward (punishment of) delinquent behavior. This underlines that new rigorously designed and controlled studies examining the effectiveness of juvenile awareness programs are needed.

A third limitation is that we were not able to find any studies on the effectiveness of programs in which ex-prisoners visit schools instead of children visiting prisons, and therefore no conclusions can be drawn about the effectiveness of such programs. Further, we could only find

studies conducted in the United States, even though juvenile awareness programs have also been implemented in several other countries. Fourth, the reported information in some primary studies was limited. For instance, three studies reported insufficient information on specific program components and were therefore not valuable in the moderator analyses that were performed for examining how individual program components are associated with program effectiveness. These analyses were therefore based on only a small number of effect sizes, resulting in a rather low statistical power and the impossibility to perform these analyses separately for the different outcomes. Consequently, there may be a true moderating effect of several program components and/or characteristics, which we were unable to detect in the present review. Finally, due to the lack of descriptive information in the primary studies, a full and proper assessment of the risk of bias in each included study could not be performed. However, to examine whether the overall effect of juvenile awareness programs was influenced by risk of bias in individual studies, the moderating effect of three study quality variables was analyzed. Specifically, we tested whether the estimated overall effect was influenced by the research design of primary studies (quasi-experimental versus RCT), sample size, and sample dropout. None of these variables were significant moderators, implying that our estimated overall effect was at least to some extent robust to bias in the results reported in the included studies.

Despite these limitations, this study provides an important contribution to our knowledge on the effectiveness of juvenile awareness programs. This review showed that juvenile awareness programs have a positive effect on attitudes toward delinquency and attitudes toward punishment. Our study also showed that the effectiveness of these programs increases as the follow-up length increases, indicating that longer follow-up periods are necessary to properly measure effects of juvenile awareness programs on delinquent behavior. It seems that most studies on the effectiveness of juvenile awareness programs that are currently available are restricted in the sense that the follow-up duration is too short for properly assessing program effects. In fact, in most studies, the

assessment was exclusively performed immediately after the program was ended. Several studies on the effectiveness of other preventive interventions also found larger positive intervention effects in later follow-up assessments than in immediate post-intervention assessments (e.g., Van der Put et al. 2017), which may be attributed to sleeper effects of interventions (Maurer et al. 2007). Sleeper effects imply that positive intervention effects—at least to some extent—need time to emerge after the interventions has ended.

This review is in particular a call for future rigorously designed randomized controlled trials on the effectiveness of juvenile awareness programs. After all, no studies have been performed in the past 25 years, despite the fact that juvenile awareness programs are still quite popular and remain in use worldwide. It is particularly important that follow-up assessments of substantial length are part of the design of future studies, so that the effectiveness of (components of) these programs can be better grasped. In addition, because of the popularity of programs in which ex-prisoners visit schools instead of children visiting prisons, it is also important that the different types of awareness programs are examined. Offering these programs may be an interesting prevention strategy, because the programs are easy to implement at low costs and may have benefits for both the juveniles and the ex-prisoners. In general, effects of preventive programs on juvenile delinquency are only small (De Vries, Hoeve, Assink, Stams, & Asscher, 2015), because such programs are aimed at juveniles of which a considerable proportion would not have started committing offenses, regardless of attending a preventive program. However, even small effects can be very meaningful in clinical practice. Future high quality research is needed to better grasp the effects of juvenile awareness programs, and to make more informative decisions on whether or not these programs should be implemented.

Conflict of interest None of the authors have any conflicts of interest to declare. This research did not involve human subjects, and therefore no consent or assent forms needed to be approved by an ethics committee.

Funding This research did not receive any grant from funding agencies in the public, commercial, or not-for-profit sectors.

References*

(*References marked with an asterisk were included in the meta-analysis)

Assink, M., & Wibbelink, C. J. M. (2016). Fitting three-level meta-analytic models in R: A step-by-step tutorial. *The Quantitative Methods for Psychology, 12*(3), 154-174. doi: 10.20982/tqmp.12.3.p154

*Berry, R. L. (1985). *Shape-Up: The effects of a prison aversion program on recidivism and family dynamics [PhD thesis]*. Greeley, CO: University of Northern Colorado.

*Buckner, J. C., & Chesney-Lind M. (1983). Dramatic cures for juvenile crime: An evaluation of a prisoner-run delinquency prevention program. *Criminal Justice & Behavior, 10*(2), 227-247.

*Cook, D. D., & Spirrison, C. L. (1992). Effects of a non-confrontational prisoner-run juvenile delinquency deterrence program. *Journal of Offender Rehabilitation, 17*(3/4), 89-99.

De Vries, H. (2017). An integrated approach for understanding health behavior; The i-change model as an example. *Psychology and Behavioral Science International Journal*, *2*(2), 1-4.

de Vries, S. L., Hoeve, M., Assink, M., Stams, G. J. J., & Asscher, J. J. (2015). Practitioner review: effective ingredients of prevention programs for youth at risk of persistent juvenile delinquency–recommendations for clinical practice. *Journal of Child Psychology and Psychiatry*, *56*(2), 108-121.

*Dean, D. G. (1982). The impact of a juvenile awareness program on select personality traits of male clients. *Journal of Offender Counseling, Services & Rehabilitation, 6*(3), 73-86.

Duval, S., & Tweedie, R. (2000a). A nonparametric 'trim and fill' method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*, 89-99.

Duval, S., & Tweedie, R. (2000b). Trim and fill: A simple funnel-plot based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*, 455-460.

Egger, M., Davey-Smith, G., & Altman, D. (2001). *Systematic reviews in healthcare*. London: British Medical Journal Books.

Farrington, D. P. (2003). Methodological quality standards for evaluation research. *The Annals of the American Academy of Political and Social Science*, *587*(1), 49-68.

Ferguson, G. A. (1966). *Statistical Analysis in Psychology & Education* (pp. 244). New York, NY: McGraw-Hill.

*Finckenauer, J. O. (1982). *Scared straight and the panacea phenomenon.* Englewood Cliffs, NJ: Prentice-Hall.

Finckenauer, J. O., Gavin, P. W., Hovland, A., & Storvoll, E. (1999). *Scared Straight!: The panacea phenomenon revisited*. Prospect Heights, NY: Waveland Press.

* GERP&DC (1979). Greater Egypt Regional Planning & Development Commission. *Menard Correctional Center: Juvenile tours impact study.* Carbondale, IL: Greater Egypt Regional Planning & Development. https://www.ncjrs.gov/pdffiles1/Photocopy/62932NCJRS.pdf

Hox, J. (2002). *Multilevel analysis: Techniques and applications.* Mahwah, NJ: Lawrence Erlbaum Associates.

Knapp, G. & Hartung, J. (2003). Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine, 22*(17)*,* 2693-2710.

*Langer, S. (1980). *Fear in the deterrence of delinquency: A critical analysis of the Rahway State Prison Lifer's program [PhD thesis].* Newark, NJ: Rutgers University.

Lebel, T. P. (2007). An examination of the impact of formerly incarcerated persons helping others. *Journal of Offender Rehabilitation*, *46*(1-2), 1-24.

*Lewis, R. V. (1983). Scared Straight – California style: Evaluation of the San Quentin SQUIRES program. *Criminal Justice & Behavior, 10*(2), 209-226.

Li, Y., Shi, L., & Roth, D. (1994). The bias of the commonly-used estimate of variance in meta-analysis. *Communications in Statistics – Theory and Methods, 23*(4), 1063-1085. doi: 10.1080/03610929408831305

Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

*Locke, T. P., Johnson, G. M., Kirigin-Ramp, K., Atwater, J. D., & Gerrard M. (1986). An evaluation of a juvenile education program in a state penitentiary. *Evaluation Review, 10*(3), 281-298.

Lundman, R. J. (1993). *Prevention and control of juvenile delinquency.* 2nd Edition. New York: Oxford University Press.

*Michigan Department of Corrections (1967). *A six month follow-up of juvenile delinquents visiting the Ionia reformatory.* Lansing, MI: Michigan Department of Corrections.

Muhammed, L. (1999). Kids and crooks revisited: Some were'Scared Straight!'. *USA Today.(April 12) D*, *4*.

Nakagawa, S., & Santos, E. S. A. (2012). Methodological issues and advances in biological meta-analysis. *Evolutionary Ecology, 26*(5), 1253-1274.

*Orchowsky, S., & Taylor, K. (1981). *The Insiders Juvenile Crime Prevention program.* Richmond, VA: Virginia Department of Corrections.

Paternoster, R. (2010). How much do we really know about criminal deterrence? *The Journal of Criminal Law and Criminology, 100*(3), 765-824.

Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2007). Performance of the trim and fill method in the presence of publication bias and between-study heterogeneity. *Statistics in Medicine, 26*(25), 4544–4562.

Petrosino, A., Turpin-Petrosino, C. & Buehler, J. (2002) "Scared Straight" and other juvenile awareness programs for preventing juvenile delinquency', *Cochrane Database of Systematic Reviews, 2.*

Petrosino, A., Turpin-Petrosino, C., Hollis-Peel, M. E., & Lavenberg, J. G. (2013). "Scared Straight" and other juvenile awareness programs for preventing juvenile delinquents. *Cochrane Database of Systematic Reviews, 4*. doi:10.1002/14651858.CD002796.pub2

Pratt, T. C., Cullen, F. T., Blevins, K. R., Daigle, L. E., & Madensen, T. D. (2006). The empirical

status of deterrence theory: A meta-analysis. In F. T. Cullen, J. P. Wright, & K. R. Blevins (Eds.), *Taking stock: The status of criminological theory* (pp. 367-396). New Brunswick, NJ: Transaction publishers.

Rosenthal, R. (1994). Parametric measures of effect size. In H. Cooper & L. V. Hedges (Eds.), *The Handbook of Research Synthesis* (pp. 239). New York, NY: Sage.

Rosenthal, R. (1995). Writing meta-analytic reviews. *Psychological Bulletin, 118*, 183-192.

Shadish, W. R., Navarro, A. M., Matt, G. E., & Phillips, G. (2000*). The effects of psychological therapies under clinically representative conditions: A meta-analysis.* Psychological Bulletin, 126(4), 512-529. doi:10.1037/0033-2909.126.4.512

Tabachnik, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Allyn and Bacon.

Terrin, N., Schmid, C. H., Lau, J., & Olkin, I. (2003). Adjusting for publication bias in the presence of heterogeneity. *Statistics in Medicine, 22*, 2113–2126.

Van Den Noortgate, W., & Onghena, P. (2003). Multilevel meta-analysis: A comparison with traditional meta-analytical procedures. *Educational and Psychological Measurement*, *63*(5), 765-790.

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods*, *45*, 576-594.

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2015). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods*, 1-21.

Van Kempen, T., Van Rooij, R., & Rovers, B. (2010). *Inzet van ervaringsdeskundigen in criminaliteitspreventie: Een verkennend onderzoek naar de mogelijkheden* [Experience expert in crime prevention: An exploratory investigation into the possibilities]. Retrieved on 21-2-2018 from https://www.btvo.nl/wordpress/wp-content/uploads/2016/06/ervaringsdeskundigen-in-criminaliteitspreventie.pdf

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36*(3), 1–48.

*Vreeland, A. D. (1981). *Evaluation of Face-to-Face: A juvenile aversion program[PhD thesis].* Dallas, TX: University of Texas.

*Yarborough, J. C. (1979). *Evaluation of JOLT as a deterrence program.* Lansing, MI: Michigan Department of Corrections.

Ziegler, S., Koch, A., & Victor, N. (2001). Deficits and remedy of the standard random effects methods in meta-analysis. *Methods of Information in Medicine, 40*(2), 148-155.

Figures and Tables

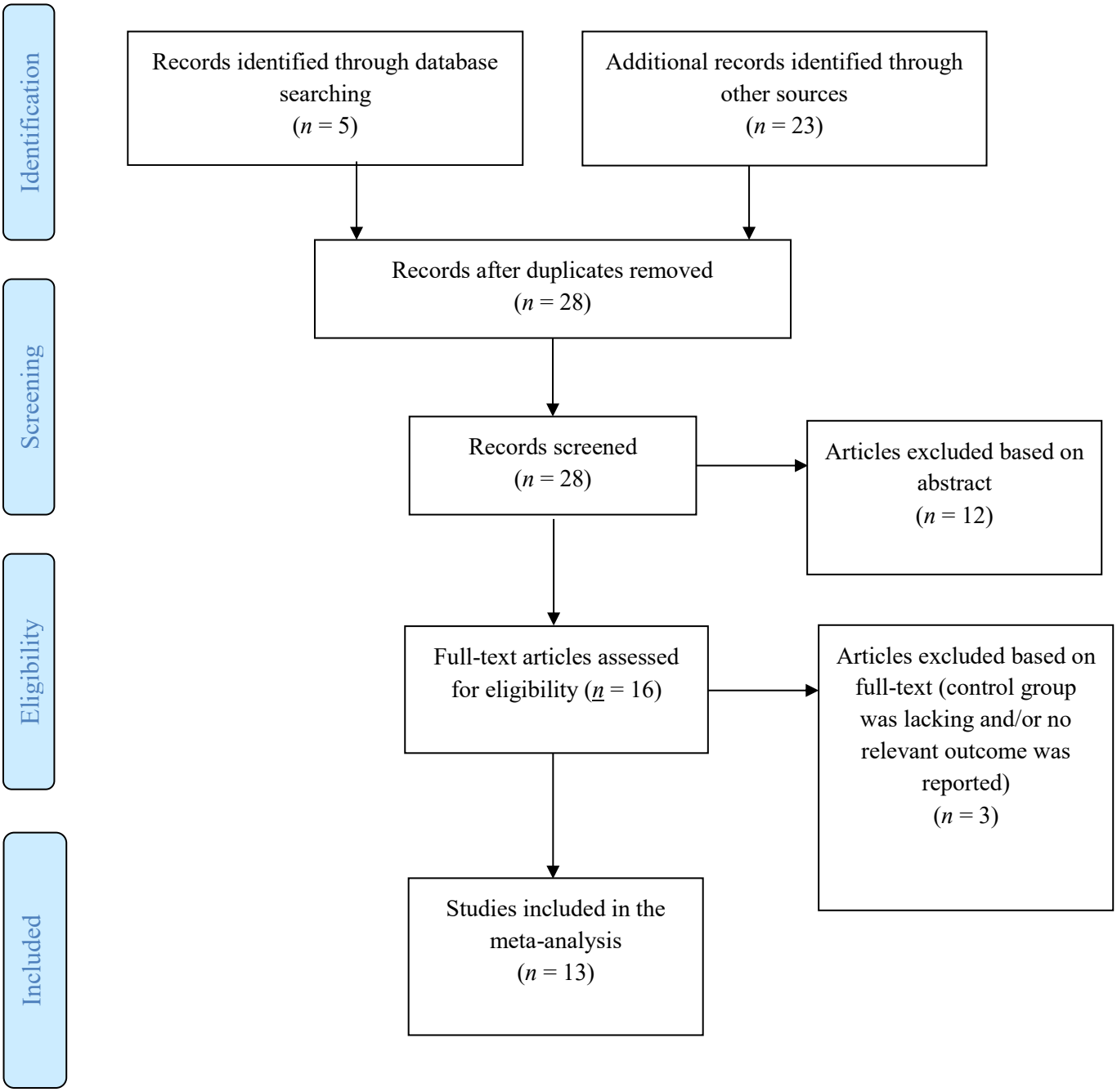

Figure 1 *Flow Chart of the Study Selection procedure*

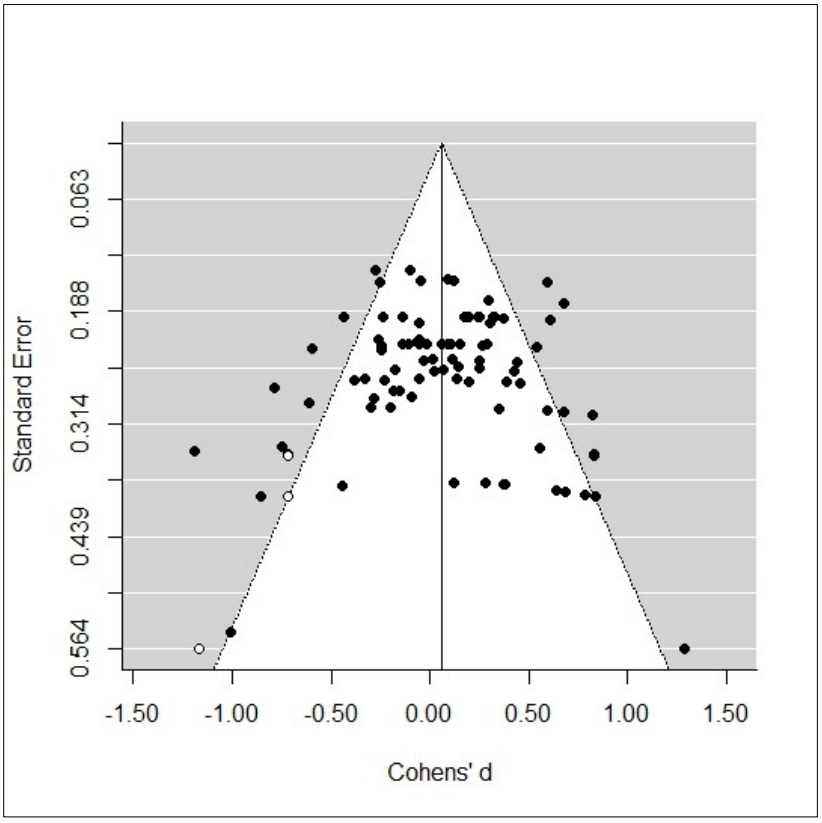

Figure 2 *Funnel Plot*

Note. A contour enhanced funnel plot is presented with the standard error on the y-axis and Cohen's *d* on the x-axis. The black dots denote the observed effect sizes, whereas the white dots denote the filled effect sizes. The solid vertical line represents the overall mean effect.

Table 1. *Characteristics of Included Studies*

Author(s), pub. year	*N*	Program name	Sample type	Sex	Outcome measurement	Type of outcome measurement	FU	Effect Size
Berry, 1985	57	SHAPE-UP	Delinquents	M	Official records	Delinquency - proportions	5	0.196
	57		Delinquents	M	Official records	Delinquency - severity	5	0.459
Buckner & Chesney-	200	Stay Straight Program	Delinquents	M	Official records	Delinquency - proportions	12	-0.104
Lind, 1983	200		Delinquents	M	Official records	Delinquency - frequency	12	-0.274
	100		Delinquents	F	Official records	Delinquency - proportions	12	0.305
	100		Delinquents	F	Official records	Delinquency - frequency	12	-0.055
Cook & Spirrison,	176	Mississippi Project Aware	Delinquents	M	Official records	Delinquency - combination	12	0.093
1992	176		Delinquents	M	Official records	School dropout	12	0.594
Dean, 1982	28	Project Aware Wisconsin	Delinquents	M	Self-report	T-JTA: Hostility	0	0.278
	28		Delinquents	M	Self-report	T-JTA: Impulsivity	0	0.636
	28		Delinquents	M	Self-report	T-JTA: Indifference	0	0.382
	28		Delinquents	M	Self-report	T-JTA: Inhibition	0	0.374
	28		Delinquents	M	Self-report	LC: Internal	0	0.836
	28		Delinquents	M	Self-report	LC: Powerful others	0	0.120
	28		Delinquents	M	Self-report	LC: Chance	0	0.782
	28		Delinquents	M	Self-report	TN Personal self concept	0	0.687
	28		Delinquents	M	Self-report	TN Family self concept	0	-0.441
	28		Delinquents	M	Self-report	TN social self concept	0	-0.853

Author(s), pub. year	*N*	Program name	Sample type	Sex	Outcome measurement	Type of outcome measurement	FU	Effect Size
Finckenauer, 1982	40	Juvenile Awareness Project	Non-delinquents	B	Official records	Delinquency - proportions	6	-1.189
	41		Delinquents	B	Official records	Delinquency - proportions	6	-0.745
	81		Both	B	Official records	Delinquency - severity	6	-0.599
	81		Both	B	Self-report	ATOL	-	-0.139
	81		Both	B	Self-report	ATPC	-	0.059
	81		Both	B	Self-report	SD: Crime	-	0.541
	81		Both	B	Self-report	SD: Law	-	-0.014
	81		Both	B	Self-report	SD: Justice	-	0.089
	81		Both	B	Self-report	SD: Prison	-	0.267
	81		Both	B	Self-report	SD: Punishment	-	0.103
	81		Both	B	Self-report	SD: I (myself)	-	-0.108
GERP&DC, 1979	53	Menard Correctional Center Tours	Non-delinquents	M	Official records	Delinquency - proportions	10	0.350
	58		Delinquents	M	Official records	Delinquency - proportions	10	-0.232
	50		Delinquents	M	Official records	Delinquency - proportions	10	-0.608
	53		Non-delinquents	M	Self-report	Jesness: Asocial Index	0	-0.298
	58		Delinquents	M	Self-report	Jesness: Asocial Index	0	0.391
	50		Delinquents	M	Self-report	Jesness: Asocial Index	0	-0.284
	53		Non-delinquents	M	Self-report	PH Self concept scale	0	-0.203
	58		Delinquents	M	Self-report	PH Self concept scale	0	-0.056
	50		Delinquents	M	Self-report	PH Self concept scale	0	-0.096
Langer, 1980	131	Juvenile Awareness Project	Delinquents	B	Official records	Delinquency - severity	10	0.293

Author(s), pub. year	*N*	Program name	Sample type	Sex	Outcome measurement	Type of outcome measurement	FU	Effect Size
	131		Delinquents	B	Official records	Delinquency - severity	22	0.678
Lewis, 1983	108	San Quentin Squires Program	Delinquents	M	Official records	Delinquency - proportions	12	-0.435
	107		Delinquents	M	Self-report	Attitudes toward crime	0	0.612
	107		Delinquents	M	Self-report	Attitudes toward police	0	0.321
	107		Delinquents	M	Self-report	Attitudes toward prison	0	0.187
	107		Delinquents	M	Self-report	Attitudes toward school	0	0.239
	107		Delinquents	M	Self-report	Attitudes toward camp	0	0.172
	107		Delinquents	M	Self-report	SD: Cell	0	0.370
	107		Delinquents	M	Self-report	SD: Crime	0	0.200
	107		Delinquents	M	Self-report	SD: Doing time	0	0.325
	107		Delinquents	M	Self-report	SD: Lock up	0	-0.235
	107		Delinquents	M	Self-report	SD: Other prisoners	0	-0.137
	107		Delinquents	M	Self-report	SD: Prison	0	0.248
Locke, Johnson, Kirigin-Ramp, Atwater, & Gerrard, 986	16	Kansas Juvenile Education Program	Delinquents	M	Self-report	Delinquency - frequency	6	1.288
	16		Delinquents	M	Self-report	Delinquency - frequency	6	-1.009
Michigan Department of Corrections, 1967	58	Ionia Reformatory Tour	Delinquents	M	Official records	Delinquency - proportions	6	-0.787
Orchowsky & Taylor,	80	The Insiders Program	Delinquents	M	Official records	Delinquency - proportions	6	-0.052

Author(s), pub. year	*N*	Program name	Sample type	Sex	Outcome measurement	Type of outcome measurement	FU	Effect Size
1981	47		Delinquents	M	Official records	Delinquency - proportions	9	0.595
	36		Delinquents	M	Official records	Delinquency - proportions	12	0.553
	80		Delinquents	M	Official records	Delinquency - frequency	6	0.152
	47		Delinquents	M	Official records	Delinquency - frequency	9	0.676
	36		Delinquents	M	Official records	Delinquency - frequency	12	0.832
	80		Delinquents	M	Official records	Delinquency - severity	6	0.287
	47		Delinquents	M	Official records	Delinquency - severity	9	0.822
	36		Delinquents	M	Official records	Delinquency - severity	12	0.833
Vreeland, 1981	79	Face to Face Program: Prison only	Delinquents	M	Official records	Delinquency - proportions	6	-0.244
	69		Delinquents	M	Self-report	Delinquency - proportions	6	0.249
	69		Delinquents	M	Self-report	ATOL	6	0.438
	69		Delinquents	M	Self-report	Having delinquent friends	6	0.111
	69		Delinquents	M	Self-report	Deterrence - certainty	6	0.015
	68		Delinquents	M	Self-report	Deterrence - severity	6	-0.029
	58		Delinquents	M	Self-report	OSIQ: Family relations	6	0.139
	58		Delinquents	M	Self-report	OSIQ: Impulsivity	6	-0.384
	59		Delinquents	M	Parent-report	Jesness: Socal competence	6	-0.329
	76	Face to Face Program: Prison + counsel	Delinquents	M	Official records	Delinquency - proportions	6	-0.244
	64		Delinquents	M	Self-report	Delinquency - proportions	6	0.252
	64		Delinquents	M	Self-report	ATOL	6	0.140
	63		Delinquents	M	Self-report	Having delinquent friends	6	0.068

Author(s), pub. year	*N*	Program name	Sample type	Sex	Outcome measurement	Type of outcome measurement	FU	Effect Size
	63		Delinquents	M	Self-report	Deterrence - certainty	6	-0.175
	63		Delinquents	M	Self-report	Deterrence - severity	6	0.428
	53		Delinquents	M	Self-report	OSIQ: Family relations	6	-0.156
	53		Delinquents	M	Self-report	OSIQ: Impulsivity	6	-0.182
	62		Delinquents	M	Parent-report	Jesness: Socal competence	6	0.023
Yarborough, 1979	169	Juvenile Offenders Learn Truth Program	Delinquents	M	Official records	Delinquency - proportions	3	-0.049
	84		Delinquents	M	Official records	Delinquency - proportions	6	-0.054
	169		Delinquents	M	Official records	Delinquency - frequency	3	-0.252
	84		Delinquents	M	Official records	Delinquency - frequency	6	-0.259
	169		Delinquents	M	Official records	Delinquency - severity	3	0.122
	82		Delinquents	M	Official records	Delinquency - severity	6	-0.066

Note. Pub. year = Publication year; *N* = total sample size; FU = Follow-up duration in months; ATPC = Attitude Toward Punishment of Criminals; SD = Semantic Differential scale; ATOL = Attitudes Toward Obeying the Law; OSIQ = Offer Self-Image Questionnaire;

Table 2 *Results of the Moderator Analyses (Bivariate Models)*

Moderator variables	# Studies	# ES	Intercept/mean *d* (95% CI)	β_1(95% CI)	*F* (df1, df2)[a]	*p*[b]	Level 2 variance	Level 3 variance
Overall Effect	13	88	0.099(-0.040, 0.238)				.042***	.039**
Study characteristics								
Year of publication	13	88	0.090(-0.035, 0.215)	0.038(0.006, 0.069)*	5.658(1, 86)	.020*	.039***	.029*
Research design					0.671(1, 86)	.415	.042***	.045**
RCT (RC)	10	80	0.069(-0.093, 0.232)					
Quasi-experimental	3	8	0.217(-0.104, 0.538)	0.148(-0.212, 0.508)				
Sample size	13	88	0.101(-0.046, 0.248)	-0.000(-0.002, 0.002)	0.014(1, 86)	.906	.043***	.043**
Dropout percentage	9	71	0.105(-0.044, 0.254)	-0.004(-0.011, 0.003)	1.545(1, 69)	.218	.050***	.028
Sample type					0.582(1, 77)	.448	.033**	.088**
Delinquent (RC)	13	75	0.064(-0.127, 0.256)					
Non-delinquent	2	4	-0.096(-0.534, 0.342)	-0.160(-0.578, 0.258)				
Sex					0.426(2, 85)	.655	.040***	.055**
Boys (RC)	11	73	0.080(-0.092, 0.252)					
Girls	1	2	0.298(-0.200, 0.796)	0.217(-0.272, 0.707)				
Both	2	13	0.147(-0.238, 0.531)	0.066(-0.355, 0.487)				
Mean age (in years)	11	77	0.107(-0.046, 0.260)	-0.036(-0.233, 0.161)	0.131(1, 75)	.718	.035***	.040**
Percentage of non-Caucasians in samples	9	61	0.128(-0.028, 0.284)	0.007(-0.001, 0.016)	2.764(1, 59)	.102	.032*	.034*
Follow-up duration	12	53	-0.0072(-0.306, 0.162)	0.042(0.008, 0.077)*	5.989(1, 51)	.018*	.016	.126***
Type of outcome measurement					7.498(2, 85)	.001**	.021*	.099***

Moderator variables	# Studies	# ES	Intercept/mean *d* (95% CI)	β_1(95% CI)	*F* (df1, df2)[a]	*p*[b]	Level 2 variance	Level 3 variance
Self-report (RC)	6	51	0.320(0.094, 0.546)**					
Official record	11	35	-0.055(-0.265, 0.154)	-0.376(-0.573, -0.178)***				
Parent-report	1	2	0.078(-0.405, 0.562)	-0.242(-0.683, 0.198)				
Type of outcome					5.515(3, 84)	.002**	.021*	.081***
Delinquency (RC)	12	38	-0.019(-0.211, 0.172)					
Delinquency risk factors	6	30	0.197(-0.022, 0.415)+	0.216(0.024, 0.407)*				
Attitudes toward delinquency	3	9	0.460(0.191, 0.730)**	0.479(0.234, 0.724)***				
Attitudes toward punishment	3	11	0.328(0.070, 0.586)*	0.347(0.115, 0.580)**				
Intervention Characteristics								
Number of contact hours	5	48	0.110(-0.044, 0.264)	-0.001(-0.014, 0.011)	0.039(1, 46)	.845	.028*	.017
Level of confrontation					1.427(2, 84)	.246	.038***	.031**
Moderate (RC)	4	45	-0.002(-0.200, 0.197)					
Low	5	20	0.194(-0.037, 0.425)+	0.195(-0.110, 0.500)				
High	3	22	0.245(-0.006, 0.495)+	0.246(-0.074, 0.566)				
Component rap session					0.346(1, 81)	.558	.034**	.039**
Absent (RC)	3	37	0.049(-0.209, 0.308)					
Present	7	46	0.142(-0.035, 0.320)	0.093(-0.221, 0.407)				
Component confinement					0.032(1, 81)	.858	.034**	.041**
Absent (RC)	5	37	0.099(-0.117, 0.315)					
Present	5	46	0.126(-0.080, 0.332)	0.027(-0.271, 0.325)				
Component facility tour					0.030(1, 81)	.862	.034**	.041**

Moderator variables	# Studies	# ES	Intercept/mean *d* (95% CI)	β_1(95% CI)	*F* (df1, df2)[a]	*p*[b]	Level 2 variance	Level 3 variance
Absent (RC)	3	31	0.132(-0.129, 0.392)					
Present	7	52	0.104(-0.077, 0.285)	-0.028(-0.345, 0.289)				
Component visitation					0.176(1, 81)	.676	.034**	.041**
Absent (RC)	6	46	0.077(-0.149, 0.303)					
Present	4	37	0.140(-0.057, 0.338)	0.063(-0.237, 0.364)				
Component dialogue					0.024(1, 81)	.877	.034**	.040**
Absent (RC)	7	62	0.106(-0.066, 0.278)					
Present	3	21	0.133(-0.159, 0.424)	0.026(-0.312, 0.365)				
Component parents involved					0.076(1, 81)	.783	.037**	.033**
Absent (RC)	9	68	0.117(-0.029, 0.263)					
Present	3	17	0.085(-0.141, 0.311)	-0.032(-0.262, 0.198)				
Component counseling					0.004(1, 81)	.952	.035**	.035**
Absent (RC)	10	74	0.111(-0.030, 0.252)					
Present	1	9	0.103(-0.194, 0.400)	-0.008(-0.287, 0.270)				
Component presentation					1.298(1, 81)	.258	.034**	.032**
Absent (RC)	8	71	0.077(-0.070, 0.225)					
Present	2	12	0.297(-0.057, 0.651)+	0.220(-0.164, 0.603)				
Intervention form					1.128(1, 86)	.291	.042***	.039**
Only group sessions (RC)	9	62	0.055(-0.107, 0.217)					
Group and individual sessions	4	26	0.223(-0.048, 0.494)	0.169(-0.147, 0.484)				

Note. # Studies = number of studies; # ES = number of effect sizes; mean *d* = mean effect size expressed in Cohen's *d*; CI = confidence interval; β_1 = estimated regression coefficient; df = degrees of freedom; [a] Omnibus test of all regression coefficients in the model.

[b] *p*-Value of the omnibus test.

$^{+}p < .1$; $^{*}p < .05$; $^{**}p < .01$; $^{***}p < .001$

Table 3 *Results of the Moderator Analysis Testing Follow-Up Length Separately for Delinquency and Attitude Outcomes*

	# Studies	# ES	Intercept (95% CI) / Mean *Z* (95% CI)	β_1 (95% CI)	*F* (df1, df2)[a]	*p*[b]
Delinquency					$F(1, 36) = 8.044$	.007**
Follow-up length: 0-6 months (RC)	7	21	-0.272 [-0.563; 0.018]+			
Follow-up length: > 6 months	6	17	0.171 [-0.126; 0.467]	0.443 [0.126; 0.759]**		
Atttitudes towards delinquency/punishment					$F(1, 18) = 5.139$	.036*
Follow-up length: 0-6 months (RC)	2	19	0.181 [0.078; 0.290]**			
Follow-up length: > 6 months	1	1	0.594 [0.227; 0.961]**	0.413 [0.030; 0.796]*		
Delinquency risk factors					$F(1, 26) = 5.057$	.033*
Follow-up length: 0-6 months (RC)	4	27	0.057 [-0.123; 0.236]			
Follow-up length: > 6 months	1	1	0.594 [0.137; 1.051]*	0.537 [0.046; 1.029]*		

Note. # Studies = number of studies; # ES = number of effect sizes; mean *d* = mean effect size expressed in Cohen's *d*; CI = confidence interval; β_1 = estimated regression coefficient; df = degrees of freedom; Level 2 variance = residual variance in effect sizes within studies; Level 3 variance = residual variance in effect sizes between studies.

[a] Omnibus test of all regression coefficients in the model.

[b] *p*-Value of the omnibus test.

+ $p < .1$; * $p < .05$; ** $p < .01$; *** $p < .001$