



UvA-DARE (Digital Academic Repository)

From antisocial to prosocial?

The effectiveness of social skills training for juvenile delinquents

van der Stouwe, T.

Publication date

2020

Document Version

Other version

License

Other

[Link to publication](#)

Citation for published version (APA):

van der Stouwe, T. (2020). *From antisocial to prosocial? The effectiveness of social skills training for juvenile delinquents*. [Thesis, fully internal, Universiteit van Amsterdam].

General rights

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

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

THE EFFECTIVENESS OF SOCIAL SKILLS TRAINING FOR JUVENILE DELINQUENTS: A META-ANALYTICAL REVIEW

A revised version of this chapter is accepted for publication as:

Van der Stouwe, T., Gubbels, J., Castenmiller, Y. L., Van der Zouwen, M., Asscher, J. J., Hoeve, M., Van der Laan, P. H., & Stams, G. J. J. M. (in press). The effectiveness of social skills training (SST) for juvenile delinquents: A meta-analytical review. *Journal of Experimental Criminology*.

ABSTRACT

OBJECTIVES

To examine how effective social skills training (SST) is in reducing delinquency for juvenile offenders and for whom and under which conditions SSTs are the most effective.

METHODS

Four multilevel meta-analyses on $K = 28$ studies, and $\#ES = 580$ effect sizes, reporting on $N = 3124$ juveniles were conducted to examine the effectiveness on juvenile offender SST on offending, externalizing problems, social skills, and internalizing problems. Moderator analyses were conducted for multiple study, sample, outcome, and treatment characteristics.

RESULTS

Small beneficial effects were only found on offending and social skills, and not for externalizing and internalizing problems. Moderator analyses showed that the effects on offending and social skills were highly dependent on the comparison group: SSTs only improved offending and social skills compared to a non-treatment control group. Effects on externalizing problems were highly dependent on outcome measure and were only found for impulsivity and when the outcome was not a self-report measure.

CONCLUSIONS

SST is possibly a too generic treatment approach to be effective in reducing juvenile delinquency because dynamic risk factors for juvenile offending are only partially targeted in SST.

INTRODUCTION

Lacking social skills has been associated with problems on various life domains, and research over the past decades has repeatedly linked a lack of social skills with juvenile delinquency and reoffending (Dishion et al., 1984; Freedman et al., 1978; Gaffney & McFall, 1981; Larson et al., 2007; Ter Laak et al., 2003). As a risk factor for delinquency, social skills are often targeted in juvenile delinquency treatment to prevent reoffending. The assumption is that reducing the social skill deficits that led to the initial delinquent behavior will reduce subsequent delinquent behavior. One of the generic program types that is therefore often applied in juvenile offender treatment is social skills training (SST) (Lipsey et al., 2010).

Research on the effectiveness of SST is elaborate, and SSTs have been included in many meta-analyses examining the effectiveness of offender treatment (Landenberger & Lipsey, 2005; Lipsey et al., 2007; Lipsey et al., 2010; Lipsey, 2009) as well as in meta-analyses on SST for emotionally and behaviorally disturbed juveniles (Ang & Hughes, 2002; Cook et al., 2008; Maag, 2006), with generally positive overall treatment effects. However, the comparison of SST within a broader denominator of offender treatment types on the one hand, or a broader target population on the other hand, leaves much unclear about the effectiveness of SST for juvenile offenders specifically. Moreover, the existing meta-analyses have had limited possibilities to determine for whom and under what circumstances SSTs are most effective for this specific target population (Kazdin, 2007; Kazdin, 2008; Kraemer et al., 2002). The present study aims to fill this gap by conducting a multi-level meta-analysis on the effectiveness of SST for juvenile delinquents on reoffending as well as other externalizing problems, social skills, and internalizing problems.

Social skills enable juveniles to adequately respond to the social environment, to deal with stressful situations, and to prevent conflicts and punishment (Libet & Lewinsohn, 1973; Matson & Wilkins, 2007). Social skills generally include multiple cognitive, affective, and behavioral processes, such as problem-solving, perspective-taking, moral reasoning, self-control, and positive behavioral skills (Ang & Hughes, 2002; Spence, 2003). Consequently, SSTs generally aim to modify social skills through addressing social interaction, pro-social behavior, and social cognitive skills (Gresham, 2002; Gresham et al., 2004; Merrell & Gimpel, 1998). Common themes in the training are: emotion recognition and dealing with emotions, active listening, giving and receiving compliments, dealing with criticism and confrontations, and resisting peer pressure (Bijstra & Nienhuis, 2003).

Although the variety in SSTs has resulted in a variety of treatment approaches, common treatment techniques are based on the following theories: social learning theory (Bandura, 1977a), operant learning theory (Skinner, 1953), social information processing (Ladd & Mize, 1983), structured learning theory (Goldstein, Sprafkin, Gershaw, & Klein, 1983), and multiple cognitive approaches (Cook et al., 2008; Kazdin, 1992). Based on these theories, treatment techniques such as modeling, positive reinforcement, coaching, and role-playing are frequently used (Maag, 2006).

Several meta-analyses have reported positive effects of SST, although few have specified treatment effects for adolescents and/or juvenile offenders specifically. For instance, SSTs have been included in a meta-analysis on treatment effectiveness for juvenile offenders aged 12 to 21, categorized as a skill building program among behavioral programs, cognitive-behavioral therapy, challenge programs, academic training, and job-related interventions (Lipsey, 2009). These skill building programs were found

to result in 12% less recidivism than a control group with a 50% recidivism rate, even when controlling for study design and demographic characteristics. Moreover, effects for these programs were larger with juveniles who were older, had a higher delinquency risk, and had a less aggressive history. Interventions were more effective with juveniles diverted to community treatment, and when the intervention implementation quality was relatively high. Although the differences between skill building program types were not significant, social skills training showed a reduction in recidivism of 13%, which was less than behavioral and cognitive-behavioral approaches, but more than challenge programs, academic training, and job-related interventions.

A mega-analysis synthesized the meta-analytic outcomes of SST meta-analyses on juveniles with emotional and behavioral problems for secondary school students (Cook et al., 2008). The included meta-analyses found overall small to large treatment effects for juveniles from 11 to 19 years old, compared to a majority of no or placebo treatment controls. Two of the included meta-analyses examined juveniles with (a risk for) antisocial behavior on broadly defined outcomes of antisocial behavior, social skills and social-cognitive skills (mean $d = .41$, Lösel & Beelmann, 2003), and social or behavioral adjustment (mean $d = .66$, Ang & Hughes, 2002). Only one meta-analysis differentiated between juveniles with internalizing and those with externalizing behavior, but found no difference in effect sizes between these groups (i.e., Beelmann, Pflingsten, & Lösel, 1994, mean $d = .45$; Cook et al., 2008). The generally positive effects for adolescents with externalizing or antisocial behavior could indicate positive effects for juvenile offenders too. However, the overall effects were based on a broad variety of outcome measures, including, but not limited to antisocial behavior and social (cognitive) skills.

While the abovementioned meta- and mega-analyses show effects that are promising for the effectiveness of juvenile offender SST, little is known about its specific effects and the conditions under which it is the most effective. Previous studies have made no or limited distinctions between offenders and juveniles with other (externalizing) behavior problems, between adolescents, children and adults, or between different outcome measures (e.g., offending). In the present meta-analytic study, we therefore only included studies examining juvenile offenders age 12 to 18, and conducted four separate meta-analyses to investigate the effects of SST on different outcomes: offending (which generally is the primary target in offender treatment), other externalizing problems (e.g., aggression), social skills, and internalizing problems. Given the promising effects in meta-analyses on juvenile offender treatment and SST, we expected positive treatment effects on all of these outcomes.

By using a multi-level meta-analysis design with the possibility to include multiple effect sizes within studies, we were able to conduct elaborate moderator analyses to shed more light on whether, for which subgroups, and under which conditions SST is effective in treating juvenile offenders. This allowed testing whether treatment effects on externalizing problems and social skills indeed led to the hypothesized effects on reoffending, which has hardly been empirically supported yet (Andrews & Dowden, 2007; Andrews & Bonta, 2010b).

Second, in contrast with two of the previously mentioned meta-analyses on SST (i.e., Ang & Hughes, 2002; Beelmann et al., 1994), we included published as well as non-published studies to reduce possible publication bias, and tested the influence of doing so by including publication status as a moderator. One existing meta-analysis (Ang & Hughes, 2002) only included studies published after 1975 to "restrict the studies to relatively contemporaneous times with regard to treatment practices, research

standards, and cultural context” (pp. 166-167). To be able to obtain as many studies (and power) as possible, we did not restrict the publication period and included quasi-experimental studies in addition to randomized studies, and included both study design characteristics as moderators.

Third, the majority of studies included in existing SST meta-analyses compared SST to a non-treatment and/or placebo control group. To examine whether SST is only effective compared to no treatment, or even superior compared to other treatment, we included the comparison condition as a moderator. Moreover, previous meta-analyses on (offender) treatment have found smaller effects for non-USA studies (Leijten et al., 2016; Van der Stouwe et al., 2014; Van Stam et al., 2014), studies conducted under clinically representative conditions (Koehler et al., 2013; Van der Stouwe et al., 2014; Weisz, Ugueto, Cheron, & Herren, 2013; Weisz et al., 2017), and of higher study quality (Moher et al., 1998), and these characteristics were included as moderators.

Fourth, the effects of gender, age, and ethnicity have been under-researched and have shown inconsistent results in previous meta-analyses on SSTs. We therefore investigated these sample characteristics as moderators. In line with outcomes of the available SST meta-analyses, we expected larger effects for older juveniles (Lipsey, 2009). Because no differential treatment effects of SST for gender and ethnicity were found in previous meta-analyses, no moderating effects for these variables were expected in the present meta-analysis. Furthermore, previous SST reviews have indicated a decrease of SST treatment effects over time (Ang & Hughes, 2002; Cook et al., 2008; Maag, 2006), which was examined by including follow-up duration as an outcome characteristic moderator.

Fifth, we included multiple treatment characteristics in moderator analyses. Given the larger SST effects for juveniles treated on diversion than for juveniles on probation and incarceration (Lipsey, 2009) we expected larger treatment effects in voluntary and non-residential treatment settings. In addition, we also tested whether higher levels of treatment integrity led to larger treatment effects as was the case in a previous meta-analysis (Goense et al., 2016; Lipsey, 2009). Moreover, according to the Risk-Needs-Responsivity principles treatment duration and intensity should be matched to the juveniles’ risk of recidivism (Andrews et al., 1990; Andrews & Dowden, 2007; Andrews & Bonta, 2010b). Finding no moderating effects for treatment duration and frequency could then be indicative of correct application of the Risk-principle. Furthermore, because we included only offenders in the meta-analysis, we were not able to test the influence of group composition (deviant-only versus individual versus mixed). Previous research has shown smaller treatment effects in deviant-only group trainings, which has been attributed to deviancy training in those groups (Ang & Hughes, 2002; Dishion et al., 1999). We could however include group size as a moderator, and we expected smaller treatment effects with larger treatment groups, hypothesizing that deviancy training would be more prevalent in larger group settings. Finally, the main treatment targets (generic social skills, problem solving, or aggression) as well as parental involvement were included as moderators to determine which SST treatment approaches show the largest treatment effects for juvenile offenders.

The following research questions were addressed in the current meta-analyses: 1) To what extent is SST effective in the prevention of recidivism? 2) To what extent is SST effective in decreasing externalizing problems, increasing social skills, and decreasing internalizing problems? 3) Which study, sample, treatment, and outcome characteristics have a moderating effect on heterogeneous outcomes? 4) What is the unique contribution of significant moderators when controlling for other significant moderating variables?

METHOD

SELECTION OF STUDIES

All studies published in English or Dutch before 2018 addressing the effectiveness of SST with juvenile offenders were included. In our search, we first set out to identify all studies on SSTs with adolescents with externalizing problem behavior, including offending. Within these studies, we then selected all studies including juvenile offenders for the present meta-analysis.

Multiple electronic databases were searched to identify relevant studies: Web of Knowledge, Scencedirect, Narcis, Ovid, Wiley, Ebscohost, Proquest, Picarta, and Google Scholar. The search string consisted of multiple elements: "skills", an intervention element ("training", "intervention", or "treatment"), an externalizing problems element ("delinquent", "externalizing", "aggression", "deviant", "conduct", "emotionally disturbed", or "problem behavior"), and a youth component ("juvenile", "youth", "adolescent", or "child"), in both English and Dutch. In addition, we searched the reference lists of related meta-analyses for relevant studies. Finally, we searched for specific SST (brand) names based on the results of the initial search, such as "social skills training", "interpersonal problem-solving skills training", "Reasoning and rehabilitation", and "Aggression replacement/regulation training".

In case studies could not be retrieved or did not report appropriate data to calculate an effect size, the authors were contacted to retrieve additional information. Only when these attempts proved to be unfruitful, the study was excluded. Figure 4 shows the flow-chart for our search.

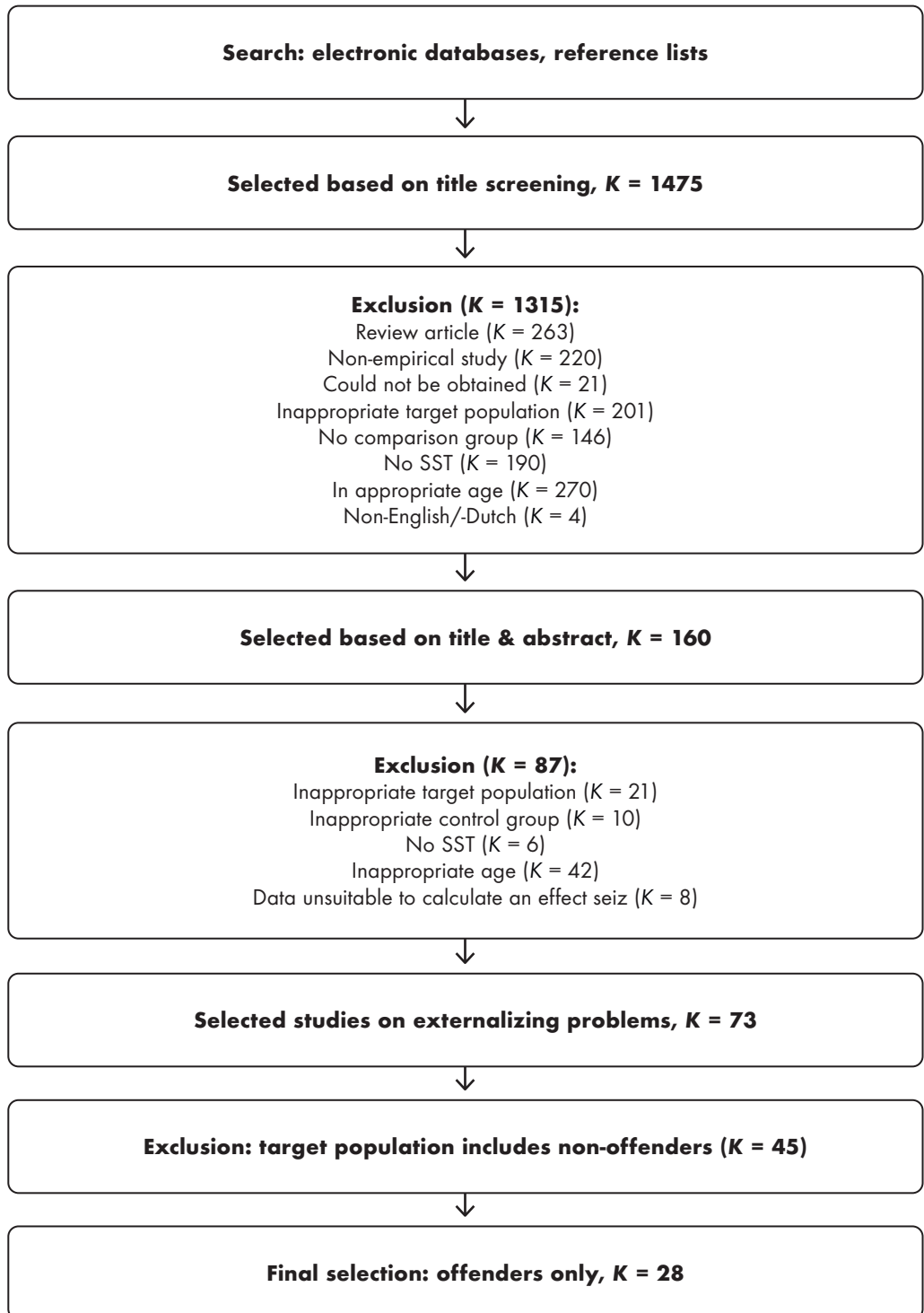
INCLUSION CRITERIA

To be included in the current meta-analyses, studies had to meet the following criteria: 1) focus on SST, defined as treatment directed at improving specific social (interactional) skills, such as social problem-solving, and assertiveness, and/or decreasing social skill deficits, 2) target juvenile offenders, or report outcomes for offenders separately, 3) target juveniles age 12 to 18, or – in case age was not reported – 7th to 12th grade, 4) employ a control group treatment design, where the control group contained juveniles from the same population, assigned to condition through random or quasi-experimental assignment, 5) report outcomes on offending, externalizing problems, social skills, and/or internalizing problems that enabled effect size calculation. Studies targeting learning disabled juveniles were excluded. The search yielded $K = 28$ studies, $\#ES = 580$, reporting on $N = 3124$ juveniles, of whom $n = 1691$ received SST treatment.

CODING THE STUDIES

Each study was coded using a detailed coding system for recording outcomes and moderators following the guideline of Lipsey and Wilson (2001). The primary outcome was offending, defined as any delinquent or illegal post-treatment activity. Secondary outcomes reported within the same study were pooled, and added as continuous moderators to include post-treatment effects on these outcomes as

Figure 4 Flow-chart for the meta-analytic search



potential moderators for offending. Secondary outcomes included externalizing problems, social skills, and internalizing problems. Externalizing problems included antisocial attitudes (e.g., cognitive distortions), impulsivity, aggression, and other externalizing problem behavior (e.g., problem behavior in the classroom, incidents, non-specific externalizing behavior). The type of problems was an outcome characteristic included as a moderator. Social skills consisted of prosocial behavior and problem solving skills, and this distinction was also included as a potential moderator. For both externalizing problems and social skills, the informant (i.e., self-report versus others report) was also included as a potential moderator.

MODERATORS

Several study, sample, and treatment characteristics were coded as potential moderators. Study characteristics were dichotomously coded for whether the study was published, was conducted outside the USA, used random (versus quasi-experimental) allocation to condition (i.e., study design), controlled for pre-existing differences between groups, was conducted in clinically representative conditions (versus efficacy studies, i.e., research design), and compared SST to alternative treatment (versus no treatment/placebo control group). Publication year and study quality were study characteristics included as continuous moderators.

For study quality, we constructed a new study quality coding list, based on the Quality Assessment Tools for Quantitative Studies (QATQS, Thomas, Ciliska, Dobbins, & Micucci, 2004), the Quality Index (QI, Downs & Black, 1998), and the Cochrane Collaboration's tool for assessing risk of bias (Higgins et al., 2011). Although all these tools have their own strengths, they also have limitations that we tried to control for with this new list. First, in a previous study we found that the QATQS did not differentiate enough in quality between studies (Van der Stouwe et al., 2014). With this list, points are only awarded for the highest study standards which most studies do not meet. The remaining points leave only little variation between normal practice less-than-perfect research. Second, the QI is a very elaborate tool that leaves relatively much room for subjective interpretation, because the criteria based on which a study meets a certain quality characteristic are not clearly defined and there is no room for studies that only partially meet a criteria. Finally, the Cochrane Collaboration Tool is deliberately qualitative in nature, which makes it unsuitable for quantitative comparison in meta-analysis.

We constructed a list of 15 items assessing publication status (one item), selection bias, study design, blinding/dependence of authors, outcome measures, attrition and dropout, intervention, and sample description (all consisting of two items, Van der Stouwe, 2016). Every item had four possible answers with the answer representing the least study quality assigned zero, and the answer representing maximum study quality assigned 3 points. Studies could therefore score between 0 and 45 points for study quality, and in the present study scores ranged from 9 to 37 points (*mean (sd) = 20.04 (6.72), median = 21*). We therefore believe that this checklist better serves the less-than-perfect research practice and the variation in study quality within those studies. The study quality list and its manual are available from the first author upon request.

Sample characteristics that were coded as potential moderators were age, and proportion of males and ethnic minority juveniles in the sample. Unfortunately, some studies provided information about grade levels instead of (average) age. To be able to include these studies in the age moderator analyses, we calculated average age based on the average age per grade level. Because there was little variance in these variables between studies, we coded all sample characteristics as dichotomous

variables: under 16 years versus 16 years and older, 75% or less males versus over 75% males, and 50% or less ethnic minority versus over 50% ethnic minority.

In addition to the outcome-specific moderators (mentioned earlier), we coded the duration of follow-up, and whether the outcome could be controlled for pretest effects as moderators. For follow-up duration, there was too little variation to be able to include follow-up as a continuous moderator. We therefore dichotomized this variable into less than 6 months follow-up versus 6 months and longer follow-up.

Several treatment characteristics were coded as potential moderators. We coded whether treatment was residential (versus non-residential), voluntary (versus compulsory), whether other treatment was provided during the SST treatment (i.e., concurrent treatment), whether treatment integrity was monitored and whether parents were involved in the training as dichotomous variables. Continuous moderators were treatment duration in weeks and treatment group size. Finally, treatment frequency was divided into three groups: unknown, once a week or less, and more than once a week. In addition, we divided the different types of SST into (general) social skills training, and programs specifically targeting problem-solving, or aggression and anger management. Studies were coded by the first, second, and third author. To determine interrater reliability, four studies were double-coded. Interrater agreement ranged from 95% to 100%.

CALCULATION AND ANALYSIS

For each study outcome we calculated an effect size of Cohen's d , using formulas from Lipsey and Wilson (2001), and Wilson (2010), with a positive effect size indicating better results for the SST group. Based on Cohen's (1988) guideline, an effect size of $d = .20$ was considered small, an effect size of $d = .50$ was considered medium, and an effect size of $d = .80$ was considered large. To control for pre-treatment differences on the outcome measure, we calculated effect sizes for both pre-treatment and post-treatment, and then subtracted the pre-treatment effect from the post-treatment effect whenever possible. When outcome effects were reported to be non-significant without reporting statistics to be able to calculate an effect size, we conservatively estimated the effect size to be zero (Lipsey & Wilson, 2001).

Several steps were taken to prepare the data for data-analysis. Effect sizes and continuous moderators were examined for outliers using their Z-distribution. Extreme values (> 3.29 SD from the mean, Tabachnick & Fidell, 2013) were winsorized by recoding them into the nearest non-outlier. For offending outcomes, one study had to be recoded, only for treatment duration in weeks (39 weeks, Barker, 1998). For externalizing behavior outcomes, one study had an outlying group size (group size = 30, Tellier, 1998), and one study had two outlying effect sizes ($d = -8.56$ and $d = -4.13$, Garrido & Sanchis, 1991) that were winsorized. Finally, for social skills outcomes, one study had a longer treatment duration (26 weeks, Leeman, Gibbs, & Fuller, 1993), and three effect sizes needed to be winsorized ($d = -1.75$ and $d = -1.69$, Ollendick & Hersen, 1979; $d = -2.38$, Scholte & Van der Ploeg, 2003; 2006). For internalizing problems, no outliers were detected. Continuous moderators were centered around their mean, categorical moderators were dummy-coded, and standard errors and sampling variance were calculated using formulas by Lipsey and Wilson (2001).

In traditional meta-analysis, effect sizes and effect size characteristics are pooled within studies, because only one effect size per study can be included in the analysis, which generally results in a loss of information and power. To retain maximum information and power, and to be able to conduct

comprehensive moderator analyses, we conducted a multi-level meta-analysis following the approach suggested by Van den Noortgate and Onghena (2003). The meta-analysis was conducted in R (version 3.4.1) with the *metafor*-package, using a 3-level random effects model to account for sampling variance (level 1), variance between effect sizes within studies (level 2), and variance between studies (level 3), which accounts for the interdependency of effect sizes that exists when multiple effect sizes per study are included (Assink & Wibbelink, 2016; Houben, Van den Noortgate, & Kuppens, 2015; Van den Bussche, Van den Noortgate, & Reynvoet, 2009; Viechtbauer, 2010). To examine heterogeneity of the effect size distribution, we tested for significant variance at level 2 and 3 using likelihood ratio tests comparing the full model to models excluding the variance parameters of level 2 and 3, respectively (Assink & Wibbelink, 2016). If there is significant variance on the two levels, the effect size distribution is considered heterogeneous, and the overall mean cannot be treated as an estimate of a common effect size. If this was the case, the model was extended by including study, sample, outcome, and treatment characteristics to examine whether those had a moderating effect on SST treatment effects. Finally, significant moderators were included in one model to examine their unique influence when controlling for other moderators in the model.

FILE DRAWER ANALYSIS

A common threat to the generalizability of meta-analytic outcomes is publication or file drawer bias (Rosenthal, 1995). Because studies with non-significant or unfavorable outcomes are published less often, studies included in meta-analysis may not be an adequate representation of all existing studies, and may therefore provide an optimistic image of actual treatment effects. We tried to control for this type of bias by including all studies we could find, and not just studies published in peer-reviewed scientific journals. In addition, we tested for funnel plot asymmetry using Egger's method (Egger, Smith, Schneider, & Minder, 1997). If no publication bias is present, the effect sizes should result in a symmetrical funnel plot (plotted against their precision), and result in a non-significant intercept in Egger's test. Furthermore, we conducted a trim and fill procedure (Duval & Tweedie, 2000a; Duval & Tweedie, 2000b) to examine the influence of (correcting for) funnel plot asymmetry using MIX 2.0 (Bax, 2011). The trim and fill procedure estimates missing effect sizes based on the existing effect size distribution. If the trim and fill procedure led to the estimation of missing effect sizes, we imputed the effect sizes within studies and reran the overall effect size analyses including these estimates.

RESULTS

The current meta-analyses consist of $K = 28$ studies, $\#ES = 580$, reporting on $N = 3124$ juveniles, of whom $n = 1691$ received SST treatment. Because not all studies reported on all examined outcome measures, the number of studies, effect sizes and juveniles differs between outcome measures. The coded study, sample, treatment and outcome characteristics are included in Appendices A – D.

OVERALL EFFECTS

Table 11 summarizes the overall effects for offending, externalizing problems, social skills, and internalizing problems.

Table 11 Results for the overall mean effect size

Outcome	K	#ES	d (SE)	95% CI	t
Offending	13	87	.12 (.04)	.05 - .19	3.22**
Externalizing problems	19	197	.22 (.14)	-.06 - .49	1.56
Social skills	18	272	.33 (.09)	.16 - .50	3.78***
Internalizing problems	6	24	-.07 (.35)	-.79 - .66	-.19

Note. K = number of studies; #ES = number of effect sizes; d (SE) = mean effect size (standard error); 95% CI = 95% Confidence Interval; t = test value for mean d difference from zero; * = $p < .05$, ** = $p < .01$, *** = $p < .001$

Offending. Offending outcomes were reported on $N = 2600$ juveniles including $n = 1410$ juveniles who received SST. A small significant overall effect was found ($d = .12$). After SST treatment, juveniles offended less than juveniles in the control group. There was significant variance between effect sizes within studies ($\sigma^2_{level2} = .01$, $\chi^2(1) = 5.03$, $p = .02$), which explained 24% of the total variance, but no significant variance between studies ($\sigma^2_{level3} = .01$, $\chi^2(1) = 3.23$, $p = .07$), which explained 14% of the total variance. However, Egger's method indicated funnel plot asymmetry ($B = .74$, $t = 2.00$, $p = .004$), and the funnel plot showed that small studies with outcomes unfavorable to SST were less often reported. After a trim and fill procedure to correct for this asymmetry (Duval & Tweedie, 2000a; 2000b), the overall effect size was no longer significant ($k = 17$, #ES = 101, $d = .02$, $t = .26$, $p = .80$).

Externalizing problems. Outcomes on other externalizing problems were reported on $N = 1034$ juveniles including $n = 531$ juveniles who received SST. No significant overall effect was found ($d = .22$). After SST treatment, juveniles showed no more or less externalizing problems than juveniles in the control group. There was significant variance between effect sizes within studies ($\sigma^2_{level2} = .21$, $\chi^2(1) = 108.43$, $p = .000$), which explained 37% of the total variance, as well as significant variance between studies ($\sigma^2_{level3} = .28$, $\chi^2(1) = 70.26$, $p = .000$), which explained 49% of the total variance. Although there was significant funnel plot asymmetry ($B = .87$, $t = 2.44$, $p = .02$), a trim and fill procedure did not indicate any missing effect sizes (Duval & Tweedie, 2000a; 2000b). The overall effect size for externalizing behavior therefore appears fairly robust to publication bias and file drawer bias.

Social skills. Outcomes on social skills were reported on $N = 900$ juveniles including $n = 459$ juveniles who received SST. A small significant overall effect was found ($d = .33$). After SST treatment, juveniles showed better social skills than juveniles in the control group. There was significant variance between effect sizes within studies ($\sigma^2_{level2} = .05$, $\chi^2(1) = 18.26$, $p = .000$), which explained 18% of the total variance, as well as significant variance between studies ($\sigma^2_{level3} = .10$, $\chi^2(1) = 75.75$, $p =$

.000), which explained 38% of the total variance. Although there was significant funnel plot asymmetry ($B = 1.24, t = 5.40, p = .000$), a trim and fill procedure did not indicate any missing effect sizes (Duval & Tweedie, 2000a; 2000b). The overall effect size for social skill therefore appears fairly robust to publication bias and file drawer bias.

Internalizing problems. Outcomes on internalizing problems were reported on $N = 209$ juveniles including $n = 118$ juveniles who received SST. No significant overall effect was found ($d = -.07$). There was no significant variance between effect sizes within studies ($\sigma^2_{level2} = .00, \chi^2(1) = 0, p = 1.000$), but there was significant variance between studies ($\sigma^2_{level3} = .65, \chi^2(1) = 17.23, p = .000$), which explained 82% of the total variance. There was no significant funnel plot asymmetry ($B = -2.25, t = -1.65, p = .112$), indicating that the overall effect size for internalizing problems is robust to publication and file drawer bias. Because internalizing problems were only reported for $K = 6$ studies and $\#ES = 24$ effect sizes, moderator analyses could not be conducted.

MODERATOR ANALYSIS

Offending. Table 12 shows the results of the moderator analyses for offending. The comparison treatment had a significant moderating effect on SST treatment effects on offending. Treatment effects were larger when SST treatment was compared to a no treatment or placebo control group, although treatment effects were significant in both comparisons. Publication year had a moderating effect as well: earlier studies reported significantly larger effects than more recent studies. The remaining study characteristics, publication status, country where the study was conducted, study design (random versus quasi-experimental condition assignment), whether outcomes were controlled for pre-test differences between groups, research setting (efficacy versus effectiveness research), and study quality, did not have a moderating effect.

For sample characteristics, no moderating effects were found. Age, the proportion of males and the proportion of ethnic minority youth in the sample did not have moderating effects on SST outcomes. Moreover, no differences were found for outcome characteristics: SST effects were similar regardless of follow-up duration, and the ability to control for pretest effects on the same measures.

Of all treatment characteristics, three significant moderating effects were found. First, treatment effects were larger when SST was provided as a voluntary treatment. The effects for obligated treatment were smaller, although still significant. Second, treatment effects were larger when juveniles received concurrent treatment in addition to the SST. With exclusive SST treatment, treatments effects were smaller, but still significant. Third, different treatment effects were found for different treatment targets. Only problem-solving training showed significant positive treatment effects, whereas generic SST and aggression training did not. No significant moderating effects were found for whether or not treatment was residential, whether treatment integrity was monitored, and whether parents were involved in treatment. The duration, group size, and frequency of treatment showed no differential treatment effects. Finally, we investigated whether effects on externalizing problems and social skills outcomes moderated effects on reoffending for the six studies ($\#ES = 26$) that reported both outcomes. No moderating effects were found for externalizing problems and social skills on the effects of SST on reoffending.

Table 12 Moderator effects for offending

Moderator	K	#ES	d/β_0	t	β_1	t_1	F
<i>Study characteristics</i>							
Published	13	87					.22
Non-USA	13	87					.03
Study design	13	87					.29
Pre-existing differences uncontrolled	13	87					.00
Research setting	13	87					.01
Compared to alternative treatment	13	87					4.64*
No treatment/placebo	6	15	.28	3.37**			
Alternative treatment	9	72	.08	2.19*	-.20	-2.15*	
Year of publication (C)	13	87	.14	4.91***	-.01	-2.39*	5.72*
Study quality (C)	13	87					.52
<i>Sample characteristics</i>							
≥ 16 years	13	87					.21
Proportion male	13	87					1.87
> 50% ethnic minority	8	68					.78
<i>Outcome characteristics</i>							
≥ 6 months follow-up	13	87					.21
Controlled for pre-test score	13	87					1.14
<i>Treatment characteristics</i>							
Residential treatment	13	87					3.06
Voluntary treatment	13	87					6.83*
Obligatory treatment	10	41	.07	2.45*			
Voluntary treatment	3	46	.19	5.10***	.12	2.61*	
Other concurrent treatment	13	87					8.70**
No concurrent treatment	11	42	.06	2.36*			
Concurrent treatment	3	45	.20	5.33***	.13	2.95**	
Monitored treatment integrity	13	87					.86
Involving parents	13	87					1.50
Treatment duration in weeks (C)	11	83					.36
Treatment group size (C)	12	86					3.83
Treatment frequency (D) ¹	13	87					.19
Treatment target (D)	13	87					5.54**
Social skills (RC)	8	27	.07	1.86			
Problem-solving	4	51	.20	5.65***	.13	2.37*	
Aggression	1	9	.04	.97	-.04	-.68	
<i>Other outcome measures</i>							
Externalizing problems	6	26					.29
Social skills	6	26					2.92

Note. K = number of studies; #ES = number of effect sizes; d/β_0 = mean effect size; t = test value for mean d difference from zero; β_1 = deviance from reference category; t_1 = test value for deviance from reference category; F = omnibus test, df: #categories-1, #ES#categories. All moderators are dichotomous unless otherwise specified. (RC) = Reference category; (C) = Continuous variable; (D) = Dummy coded; ¹ unknown, ≤ once a week, > once a week. * = $p < .05$, ** = $p < .01$, *** = $p < .001$

Externalizing problems. Moderator analyses for externalizing problems were conducted for the same moderators as for offending. In addition, the moderating effects of the type of externalizing behavior (impulsivity, antisocial attitudes, aggression and other externalizing behavior) and informant (self versus others) were examined. Table 13 shows only moderators that had a significant moderating effect. The moderator analyses showed that two study characteristics significantly influenced SST treatment effects on externalizing problems. The country where the research was conducted, and whether the comparison group received treatment differentiated between treatment effects. SST showed positive treatment effects for studies conducted in the USA, and when the comparison group received no alternative treatment or a placebo. The remaining study characteristics, publication status, study design, controlling for pre-existing differences, research setting, year of publication and study quality did not have a moderating effect. Sample characteristics had no moderating influence on SST effects on externalizing problems.

Two outcome characteristics moderated treatment effects on externalizing problems. The type of informant had a significant moderating effect. SST only showed treatment effects when others (parents, teachers, or staff members) reported about externalizing problems, and not when the measures came from self-report. In addition, the type of externalizing problems measured influenced SST treatment effects. Significant treatment effects were only found for impulsivity, while treatment effects on antisocial attitudes, aggression and externalizing behaviors not otherwise specified were not significant. Follow-up duration and controlling for pre-test scores of the outcome measure did not influence SST treatment effects.

Of all treatment characteristics, only voluntary treatment showed different treatment effects. Only when the SST was voluntary (and not compulsory), there were significant treatment effects. Treatment effects did not differ in residential settings, with other concurrent treatment, when treatment integrity was monitored, when parents were involved, or depending on duration, group size, frequency, and treatment target.

Social skills. Moderator analyses for social skills were conducted for the same moderators as for offending. In addition, the moderating effects of the type of social skills (prosocial behavior versus problem-solving skills) and informant (self versus others) were examined. Only for follow-up duration, moderating effects could not be examined, because all studies reporting on social skills measured outcomes within 6 months after treatment termination. Again, only significant moderators are depicted in Table 13. Of all study characteristics, only the comparison treatment had a significant moderating effect on SST treatment effects on social skills. SST only showed significant treatment effects compared to a no treatment or placebo control group and not compared to another comparison treatment. Publication status, country where the study was conducted, study design, controlling for pre-existing differences between groups, research setting, publication year and study quality did not have moderating effects on social skills. Moreover, sample characteristics (gender, age, and ethnicity) and outcome characteristics (controlling for pre-test score, whether prosocial skills or problem-solving was measured, and informant) did not have a moderating effect either.

Three treatment characteristics had a significant moderating effect. First, SST only showed treatment effects in residential settings, while none-residential SSTs showed no significant treatment effects. Moreover, SST only had treatment effects on social skills when treatment integrity was not monitored or

Table 13 Significant moderator effects for externalizing problems and social skills

Moderator	K	#ES	d/β_0	t	β_1	t_1	F
Externalizing problems							
<i>Study characteristics</i>							
Non-USA	19	197					4.11*
USA	13	147	.41	2.59*			
Non-USA	6	50	-.14	-.63	-.54	-2.03*	
Compared to alternative treatment	19	197					3.99*
No treatment/placebo			.35	2.25*			
Alternative treatment			.04	.25	-.31	-2.00*	
<i>Outcome characteristics</i>							
Informant	19	197					11.37***
Other report	13	82	.41	2.49*			
Self-report	12	115	.02	.09	-.39	3.37***	
Outcome type (D)	19	197					5.40***
Externalizing behavior (RC)	10	51	.12	.59			
Impulsivity	6	15	.86	3.85***	.75	2.73*	
Antisocial attitudes	8	100	.15	.89	.04	.15	
Aggression	8	31	.27	1.46	.16	.63	
<i>Treatment characteristics</i>							
Voluntary treatment	19	197					4.62*
Obligatory treatment	13	95	.03	.20			
Voluntary treatment	6	102	.61	2.76**	.58	2.15*	
Social Skills							
<i>Study characteristics</i>							
Compared to alternative treatment	18	272					29.07***
No treatment/placebo	12	140	.51	6.11***			
Alternative treatment	11	132	.15	1.73	-.37	-5.39***	
<i>Treatment characteristics</i>							
Residential treatment	18	272					4.16*
Non-residential	5	58	.10	.72			
Residential	13	214	.44	4.69***	.34	2.04*	
Monitored treatment integrity	18	272					5.12*
Unmonitored	14	135	.44	4.87***			
Monitored	4	137	.04	.26	-.40	-2.26*	
Treatment target (D)	18	272					3.15*
Social skills (RC)	11	220	.30	3.32**			
Problem-solving	2	29	.84	3.71***	.54	2.19*	
Aggression	5	23	.15	.85	-.16	-.82	

Note. K = number of studies; #ES = number of effect sizes; d/β_0 = mean effect size; t = test value for mean d difference from zero; β_1 = deviance from reference category; t_1 = test value for deviance from reference category; F = omnibus test, df : #categories-1, #ES-#categories. All moderators are dichotomous unless otherwise specified. (RC) = Reference category; (D) = Dummy coded; * = $p < .05$, ** = $p < .01$, *** = $p < .001$

mentioned, while studies that did report about monitoring of treatment execution showed no treatment effects. Finally, problem-solving skills training showed the largest effects on social skills, followed by generic social skills training, but training targeting aggression did not show any treatment effects on social skills. No moderating effects were found for voluntary treatment, other concurrent treatment, involving parents in treatment, and treatment duration, group size, or frequency.

MODERATOR ANALYSIS WITH MULTIPLE MODERATORS

Offending. The results of the multivariate analyses can be found in Table 14. For offending, only one moderator retained significant moderating effect when controlling for other significant moderators. SST only showed significantly larger treatment effects when it was compared to a no treatment/placebo control group than when it was compared to alternative treatment. The other significant moderators: publication year, whether treatment was voluntary, whether other treatment was concurrent and whether problem-solving was the main treatment target, did not retain their moderating effects.

Externalizing problems. As Table 14 shows, two of the five significant moderators showed unique moderating effects when controlling for other significant moderators. SST treatment effects on externalizing problems were only found when the outcome measure was not self-reported and when impulsivity was measured. The country where the study was conducted, the comparison treatment and whether the treatment was offered voluntarily did not retain their moderating effects.

Social skills. For SST effects on social skills outcomes, only one moderator retained its moderating effects when controlling for other significant moderators. SST only showed treatment effects on social skills when it was compared to a no treatment/placebo control group, and not when it was compared to alternative treatment. Whether the treatment was residential, whether treatment integrity was monitored and the specific treatment target did not retain their moderating effects.

DISCUSSION

A series of multi-level meta-analyses were conducted to examine the effectiveness of SST for juvenile offenders on offending, externalizing problems, social skills, and internalizing problems. In line with previous studies, small treatment effects were found for offending and social skills (Cook et al., 2008; Lipsey, 2009), but no treatment effects were found for externalizing and internalizing problems. A trim and fill procedure showed that it is unlikely that the overall effect sizes for externalizing problems, social skills, and internalizing problems were affected by publication and file-drawer bias. However, a trim and fill procedure for offending resulted in a non-significant overall effect, indicating that studies with negative treatment effects are less likely to be reported, and that the available research base may overestimate the actual effects of SST on juvenile (re)offending.

Moderator analyses revealed that, for offending, larger effects were found for older studies, when SST was compared to a no treatment/placebo control group, when treatment was voluntary, when other treatment was concurrent, or when the treatment targeted problem-solving skills specifically. However, when controlling for other significant moderators, only the comparison treatment retained its

Table 14 Multivariate models with significant moderators for offending, externalizing problems and social skills

Outcome	d/β_0	t	β_1	t_1	$(df, df) F$
Offending (K = 13, #ES = 87)	.26	2.60*			(5, 81) 3.34**
<i>Study characteristics</i>					
Year of publication (C)			.00	.09	
Comparison: alternative treatment			-.22	-2.08*	
<i>Treatment characteristics</i>					
Voluntary treatment			-.08	-.72	
Other concurrent treatment			.19	1.39	
Target: Problem-solving			.04	.24	
Externalizing problems (K = 19, #ES = 197)	.45	1.88			(5, 191) 6.07***
<i>Study characteristics</i>					
Non-USA			-.36	-1.15	
Comparison: alternative treatment			-.29	-1.94	
<i>Outcome characteristics</i>					
Informant: self-report			-.28	-2.41*	
Outcome: impulsivity			.54	3.04**	
<i>Treatment characteristics</i>					
Voluntary treatment			.36	1.14	
Social skills (K = 18, #ES = 272)	.25	1.19			(5, 266) 7.38***
<i>Study characteristics</i>					
Comparison: alternative treatment			-.33	-4.72***	
<i>Treatment characteristics</i>					
Residential treatment			.13	.81	
Monitored treatment integrity			-.19	-1.33	
Target: Social skills			.23	1.27	
Target: Problem-Solving			.47	1.80	

Note. d/β_0 = mean effect size; t = test value for mean d difference from zero; β_1 = deviance from reference category; t_1 = test value for deviance from reference category; F = omnibus test; K = Number of studies; #ES = Number of effect sizes; (C) = Continuous variable; * = $p < .05$, ** = $p < .01$, *** = $p < .001$

significant moderating effect. Consequently, although SST is better than doing nothing in the prevention of juvenile (re)offending, its superiority over treatment alternatives is questionable. Interestingly, no moderating effects were found for follow-up duration, indicating that SST effects are generally consistent over time, even though previous meta-analyses have reported otherwise (Ang & Hughes, 2002; Cook et al., 2008; Maag, 2006).

For externalizing problems, SST only showed treatment effects in the USA, when compared to a no treatment/placebo control group, when the outcome was not a self-report measure, when impulsivity was the outcome measure, and when the treatment was voluntary. Multivariate analysis showed that only outcome characteristics had a unique positive effect on youth outcomes. Treatment effects for externalizing behavior were only significant for impulsivity. A possible explanation could be that impulsivity is comprised of the least underlying construct subfactors (see e.g., Sharma, Markon, & Clark, 2014), compared to externalizing behavior, antisocial attitudes, and aggression, which could make it easier to objectively define and rate. Consequently, less variance within this outcome may enable easier detection of significant effects. The broad band constructs of externalizing behavior, antisocial attitudes, and aggression may therefore be less valid. However, the lack of significant treatment effects on antisocial attitudes is unexpected. Given the fact that SSTs are generally cognitive-behavioral in nature (Cook et al., 2008), we would surely expect significant treatment effects for this more cognitively-oriented outcome type.

SST treatment effects on externalizing problems were only significant when the outcomes were not reported through self-report, but by parents, teacher, or SST trainers, which is in line with literature showing that juvenile delinquents generally underreport their behavioral problems (Breuk, Clauser, Stams, Slot, & Doreleijers, 2007; Vreugdenhil, Van den Brink, Ferdinand, & Wouters, 2006). It could, however, also indicate that others overestimate the behavioral problems as opposed to juveniles underestimating them, while low agreement between informants is not uncommon for externalizing problems with juvenile offenders (De Los Reyes et al., 2015; Forehand, Frame, Wierson, Armistead, & Kempton, 1991).

Finally, significant treatment effects on social skills were only found when comparing SST to a no treatment control group, when treatment was residential, when integrity was unmonitored, and for treatment targeting generic social skills or problem-solving capabilities. In the multivariate analyses, and similar to offending outcomes, only the use of a non-treatment comparison group yielded a significant effect for SST. This would indicate that SST is indeed successful in improving social skills, but that it is not superior to alternative treatment in doing so. Arguably, other (cheaper) treatment alternatives would suffice just as much.

Interestingly, no moderating effects were found for externalizing problems and social skills (dynamic risk factors) on offending. Treating juvenile offenders with SST assumes that juvenile offenders have less social skills than non-delinquent juveniles (Dishion et al., 1984; Freedman et al., 1978; Gaffney & McFall, 1981; Larson et al., 2007; Ter Laak et al., 2003). Arguably, lack of social skills is considered to be a cause of delinquent behavior. Thus, improving social skills (with SST) would lead to a reduction of delinquent behavior. For the present meta-analysis, we therefore expected that larger treatment effects on externalizing problems and social skills would result in significantly larger treatment effects on offending, which was not the case. Unfortunately, only a small number of the included studies reported on social skills, externalizing problems as well as offending, and the lack of moderating effect could be

due to a lack of statistical power. Given the present test statistics, this explanation is only plausible for social skills ($\beta_1 = .19, t_1 = 1.71, p = .10$), and not for externalizing behavior ($\beta_1 = .08, t_1 = .54, p = .60$).

The lack of SST treatment effects for juvenile offenders could indicate that targeting social skills as a main risk factor for delinquency might be outdated. First, it is almost never included as a separate risk factor based on contemporary risk assessment research (see e.g., Assink et al., 2015; or Jolliffe, Farrington, Piquero, Loeber, & Hill, 2017). Second, while there is little empirical evidence supporting the relative importance of different risk factors (Singh & Fazel, 2010), social skills deficits are not included as one of the *Central Eight* most important risk factors for reoffending (Andrews & Bonta, 2010a). At best, some overlap could be considered with the risk factor *antisocial cognition*, and most SSTs may indirectly focus on the *Moderate Four* risk factors (i.e., family/marital circumstances, school/work, leisure/recreation, substance abuse). Third, the dynamic predictive validity of social skills deficits is questionable, given the fact that a recent study found that only changes in antisocial attitudes/behaviors and aggression specifically, and not changes in social skills, were predictive of a recidivism reduction for juvenile offenders after residential placement, although all three constructs (regardless of change) were predictive of recidivism (Baglivio, Wolff, Jackowski, & Greenwald, 2017). Juvenile offender treatment should therefore target risk factors such as antisocial attitudes and aggression more specifically than SST does.

Moreover, the fact that SST shows similar effects as any alternative treatment might support the *dodo-bird hypothesis*, that is, the assumption that all treatments will be equally effective based on their common therapeutic characteristics (Wampold et al., 1997). This should not be too surprising given the fact that multiple social skills should be addressed and modeled in a therapeutic relation alone. However, recent reviews have shown that – in contrast to the *dodo-bird hypothesis* – most treatments still show better effects on their primary treatment target than alternative treatment at posttest, but not at follow-up (Marcus, O’Connell, Norris, & Sawaqdeh, 2014; Weisz et al., 2017). Given the lack of treatment effects on social skills and offending (i.e., the primary outcomes) when compared to alternative treatment, SST would then fair worse than other treatments at post treatment, at least in improving social skills and reducing reoffending for juvenile offenders.

This study needs to be interpreted in light of some limitations. First, as is the case with every meta-analytic study, we had to depend on the quality and elaborateness of reporting in the included studies. The lack of (explicitly) reporting about characteristics such as age, ethnicity, follow-up duration, and treatment (techniques) has therefore limited the possibilities for moderator analyses in particular. Furthermore, only a small number of studies reported outcomes about internalizing problems, and moderator analyses could therefore not be conducted for this outcome. Only 6 studies reported on externalizing problems or social skills in addition to offending outcomes, which has limited the power of moderator analyses including these outcomes. Third, several studies were excluded because they could not be obtained ($K = 21$), mostly because they were too old to be available (digitally), or did not report data suitable to calculate an effect size ($K = 8$).

To our knowledge, the present study is the first to examine SST effects for adolescent juvenile offenders. In contrast to existing meta-analyses we conducted separate multi-level meta-analyses for four separate outcomes: offending, externalizing problems, social skills, and internalizing problems. Although SST showed small positive effects on offending and social skills, the moderator analyses showed that

these effects were highly dependent on the type of comparison group: SSTs only improved social skills compared to a non-treatment control group. Significant effects on externalizing problems were highly dependent on outcome measure, and were only found for impulsivity and when the outcome was not a self-report measure. SST is possibly a too generic treatment approach to be effective in reducing juvenile delinquency because dynamic risk factors for juvenile offending are only partially targeted in SST.

APPENDICES

APPENDIX A

Study and sample characteristics of the studies included in the meta-analysis

No	Authors	Study					Sample					
		pub	USA	study	diff	research	ctrl (#ES)	Year	qual	Age	% male (#ES)	% min.
1	Jeong, Fenoiff, & Martin	yes	yes	QE	no	effect	yes	2017	25	<16	>75	>50
2	Bunford	no	yes	random	yes	efficacy	no	2016	21	≥16	>75	≤50
3	Kaya & Buzlu	yes	no	QE	yes	effect	no	2016	12	≥16	>75	-
4	Van der Stouwe, Asscher, Stams, Hoewe, & Van der Laan	no	no	QE	yes	effect	yes	2015	37	<16	≤75	>50
5	Erickson	no	yes	QE	no	effect	yes	2013	23	≥16	≤75	≤50
6	Latzman	no	yes	QE	no	effect	no	2008	9	≥16	>75	>50
7	Barnoski & Aos	no	yes	QE	yes	effect	yes	2004	18	<16	>75	-
8	Mitchell & Palmer	yes	no	QE	yes	effect	no	2004	12	≥16	>75	-
9	Ang	yes	no	QE	yes	effect	no	2003	15	<16	≤75	>50
10	Scholte & Van der Ploeg	yes	no	QE	yes	effect	no (4), yes (12)	2003	18	<16	≤75	≤50
11	Barker	no	yes	QE	yes	effect	no	1998	10	<16	≤75	>50
12	Teller	no	yes	QE	yes	efficacy	yes	1998	18	≥16	>75	>50
13	Pullen	no	yes	random	no	effect	yes	1996	20	≥16	>75	≤50
14	Leeman et al.	yes	yes	random	yes	effect	no (5), yes (7)	1993	27	≥16	>75	≤50
15	Elrod & Minor	yes	yes	QE	no	effect	yes	1992	18	<16	>75	>50
16	Garrido & Sanchis	yes	no	QE	no	efficacy	no	1991	13	<16	>75	-
17	Steele	no	yes	QE	no	effect	no	1991	10	<16	≤75	≤50
18	Guerra & Slaby	yes	yes	random	yes	efficacy	no	1990	21	≥16	≤75 (60), >75 (28)	-
19	Shivritan	yes	no	random	no	effect	no (67), yes (67)	1988	17	≥16	>75	-
20	Botcher	no	yes	QE	yes	effect	yes	1985	26	<16	≤75	>50
21	Long & Sherer	yes	yes	QE	yes	efficacy	yes	1985	29	<16	>75	>50
22	Bowman & Auerbach	yes	yes	random	yes	efficacy	no	1982	26	<16	>75	≤50
23	Schlichter & Horan	yes	yes	random	yes	efficacy	no (11), yes (11)	1981	26	≥16	>75	-
24	Toffe-Tips	no	yes	QE	yes	efficacy	no	1980	27	≥16	≤75	-
25	Feindler	no	yes	QE	yes	effect	no	1979	27	<16	≤75	≤50
26	Green-Burns	no	yes	random	no	effect	yes	1979	13	<16	>75	≤50
27	Ollendick & Hersen	yes	yes	QE	yes	effect	yes	1979	21	<16	>75	≤50
28	Sarason & Ganzler	yes	yes	QE	no	efficacy	no (3), yes (3)	1973	22	≥16	>75	-

Note. No = study number; pub = published yes/no; USA = conducted in the USA yes/no; study = allocation to condition: random or quasi-experimental (QE); diff = controlled for pre-existing differences between groups; yes/no; research = research design: efficacy or effectiveness study; ctrl (#ES) = control group received alternative treatment yes/no (number of effect sizes); qual = study quality score; % male (#ES) = proportion males (number of effect sizes); % min. = proportion ethnic minority.

APPENDIX B

Treatment characteristics and outcome characteristics for offending of the studies included in the meta-analysis

No	Authors	year	res	vol	Treatment				Offending								
					oth (#ES)	integ	par	dur	group	freq	targ	n	FU (#ES)	pre	pd (#ES)		
1	Jeong et al.	2017	no	no	no	no	yes	12	-	-	-	SS	535	-	≥6	no	.01 (1)
2	Bunford	2016	yes	yes	no	no	no	7	6	-	-	SS	-	-	-	-	-
3	Kayo & Buzlu	2016	yes	yes	no	no	no	10	11	>1	>1	AA	-	-	-	-	-
4	Van der Stouwe et al.	2015	no	no	no	no	yes	11	1	1	1	SS	222	≥6	-	no	.06 (10)
5	Erickson	2013	yes	yes	yes	yes	no	10	10	>1	>1	AA	-	-	-	-	-
6	Latzman	2008	yes	yes	no	no	no	9	14	>1	>1	SS	-	-	-	-	-
7	Barnoski & Aos	2004	yes	no	no	no	no	10	10	>1	>1	AA	1229	<6	-	no	.03 (9)
8	Mitchell & Palmer	2004	yes	yes	no	no	no	-	9	U	U	SS	62	<6	-	no	.15 (2)
9	Ang	2003	no	no	no	no	no	-	8	U	U	PS	-	-	-	-	-
10	Scholte & Van der Ploeg	2003	yes	no	no	no	no	-	10	U	U	SS	-	-	-	-	-
11	Barker	1998	no	no	no	no	no	130	7	>1	>1	PS	14	<6	-	no	.17 (1)
12	Tellier	1998	yes	no	yes	no	yes	12	30	1	1	SS	-	-	-	-	-
13	Pullen	1996	no	no	no	yes	no	18	7	U	U	SS	40	<6 (3), ≥6 (3)	-	yes	-.06 (6)
14	Leeman et al.	1993	yes	no	no	no	no	26	20	>1	>1	SS	57	≥6 (2)	-	yes	.47 (2)
15	Elrod & Minor	1992	no	no	yes	no	yes	8	7.5	1	1	SS	43	<6	-	no	.01 (2)
16	Garrido & Sanchis	1991	yes	no	no	no	no	12	14	1	1	SS	-	-	-	-	-
17	Steele	1991	yes	no	no (3), yes (3)	no	no	6	18	U	U	PS	72	<6 (2), ≥6 (4)	yes (2), no (4)	.34 (6)	
18	Guerra & Slaby	1990	yes	yes	no	no	no	12	12	1	1	PS	57	≥6	no	.17 (4)	
19	Shivritan	1988	yes	no	no	yes	no	-	1	U	U	SS	28	<6	no	.47 (2)	
20	Botcher	1985	yes	yes	yes	no	no	15	12.5	>1	>1	PS	28	≥6	yes (32), no (8)	.20 (40)	
21	Long & Sherer	1985	no	yes	no	no	no	2	9	1	1	SS	-	-	-	-	-
22	Bowman & Auerbach	1982	yes	yes	no	no	no	-	1	U	U	PS	-	-	-	-	-
23	Schlichter & Horan	1981	yes	yes	yes	no	no	5	1	>1	>1	AA	-	-	-	-	-
24	Toffe-Tippis	1980	no	no	no	no	no	3	10	>1	>1	SS	-	-	-	-	-
25	Feindler	1979	yes	yes	no	no	yes	10	1	>1	>1	AA	-	-	-	-	-
26	Green-Burns	1979	no	no	no	no	no	8	1	U	U	AA	-	-	-	-	-
27	Ollendick & Hersen	1979	yes	no	yes	no	no	10	1	1	1	SS	-	-	-	-	-
28	Sarason & Ganzer	1973	yes	no	no	no	no	4	4.5	>1	>1	SS	128	<6	no	.11 (2)	

Note. No = studynumber; res = treatment was residential yes/no; vol = treatment was voluntary yes/no; oth (#ES) = concurrent treatment during the SST treatment yes/no (number of effect sizes); integ = treatment integrity was monitored yes/no; par = parental involvement in training; dur = treatment duration in weeks; group = treatment group size; freq = treatment frequency is U(unknown), 1, or more times per week; targ = treatment target is social skills (SS), problem-solving (PS) or aggression and anger management (AA); n = sample size; FU (#ES) = follow-up duration in months (number of effect sizes); pre = controlled for pre-test score yes/no; pd (#ES) = mean effect size (number of effect sizes).

APPENDIX C

Outcome characteristics for externalizing outcomes of the studies included in the meta-analysis

No	Authors	Year	n	FU (#ES)	pre	Externalizing			Construct d (#ES)		
						pd (#ES)	inf (#ES)	EXT	IMP	ANT	AGG
1	Jeong et al.	2017	-	<6	-	-	-	-	-	-	-
2	Bunford	2016	12	<6	yes	-1.07 (6)	S (3), O (3)	-.96 (2)	-1.15 (1)	-	1.13 (3)
3	Kaya & Buzlu	2016	65	<6	yes	.36 (9)	S	-	-	.33 (5)	.39 (4)
4	Van der Stouwe et al.	2015	223	<6	yes	.25 (13)	S	-	.37 (1)	.24 (12)	-
5	Erickson	2013	60	<6	yes	.11 (3)	O	-.00 (1)	-	-	.16 (2)
6	Latzman	2008	-	-	-	-	-	-	-	-	-
7	Barnoski & Aos	2004	-	-	-	-	-	-	-	-	-
8	Mitchell & Palmer	2004	-	-	-	-	-	-	-	-	-
9	Ang	2003	105	<6	yes	.14 (6)	S	-	-	.00 (2)	.21 (4)
10	Scholle & Van der Ploeg	2003	192	<6 (4), ≥6 (4)	no	-.56 (8)	O	-.56 (8)	-	-	-
11	Barker	1998	-	-	-	-	-	-	-	-	-
12	Teller	1998	72	<6	yes	.14 (2)	S	-	-	.14 (2)	-
13	Pullen	1996	31	<6	yes	.50 (10)	S	-	.46 (1)	.50 (9)	-
14	Leeman et al.	1993	39	<6	yes	.86 (6)	S (2), O (4)	.86 (6)	-	-	-
15	Elrod & Minor	1992	-	-	-	-	-	-	-	-	-
16	Garrido & Sanchis	1991	20	<6	yes	-6.35 (2)	O	-6.35 (2)	-	-	-
17	Steele	1991	-	-	-	-	-	-	-	-	-
18	Guerra & Slaby	1990	80	<6	no	1.06 (60)	S (42), O (18)	-	1.94 (6)	.65 (48)	1.04 (6)
19	Shivratton	1988	29	<6	yes	-.20 (12)	S	-	-	-.20 (12)	-
20	Botcher	1985	-	-	-	-	-	-	-	-	-
21	Long & Sherer	1985	-	-	-	-	-	-	-	-	-
22	Bowman & Auerbach	1982	10	<6	yes	.77 (7)	S (2), O (5)	.49 (3)	.98 (4)	-	-
23	Schlichter & Horan	1981	19	<6	yes	-.52 (20)	S (12), O (8)	-	.21 (10)	.62 (10)	-
24	Toffe-Tippis	1980	20	<6	yes	.60 (24)	O	.60 (24)	-	-	-
25	Feindler	1979	8	<6	yes	1.49 (3)	S (2), O (1)	-	2.10 (2)	-	.26 (1)
26	Green-Burns	1979	10	<6	no	.02 (2)	O	.04 (1)	-	-	.01 (1)
27	Ollendick & Hersen	1979	18	<6	no	.75 (2)	O	.75 (2)	-	-	-
28	Sarason & Ganzer	1973	128	<6	no	-.10 (2)	O	-.10 (2)	-	-	-

Note. n = sample size; FU (#ES) = follow-up duration in months (number of effect sizes); pre = controlled for pre-test score yes/no; pd (#ES) = mean effect size (number of effect sizes); inf (#ES) = informant is self (S) or others (O) (number of effect sizes); Construct d (#ES) = average effect size (number of effect sizes) measuring externalizing behavior (EXT), impulsivity (IMP), antisocial attitudes (ATT), and aggression (AGG).

APPENDIX D

Outcome characteristics for social skills and internalizing outcomes of the studies included in the meta-analysis

No	Authors	Year	n	FU	pre (#ES)	Social skills		Internalizing							
						pd (#ES)	inf (#ES)	Construct d (#ES)	PSS	n	FU	pre	pd (#ES)		
1	Jeong et al.	2017	-	-	-	-	-	-	-	-	-	-	-	-	-
2	Bunford	2016	12	<6	yes	.71 (4)	O	.71 (4)	-	12	<6	yes	-.72 (1)	-	-
3	Kaya & Buzlu	2016	65	<6	yes	.54 (2)	S	-	.54 (2)	-	-	-	-	-	-
4	Van der Stouwe et al.	2015	223	<6	yes	-.08 (11)	S	-.05 (9)	-.24 (2)	-	-	-	-	-	-
5	Erickson	2013	60	<6	yes	.02 (1)	O	.02 (1)	-	-	-	-	-	-	-
6	Latzman	2008	35	<6	yes	.37 (18)	S	.33 (15)	.56 (3)	-	-	-	-	-	-
7	Barnoski & Aos	2004	-	-	-	-	-	-	-	-	-	-	-	-	-
8	Mitchell & Palmer	2004	-	-	-	-	-	-	-	-	-	-	-	-	-
9	Ang	2003	-	-	-	-	-	-	-	-	-	-	-	-	-
10	Scholte & Van der Ploeg	2003	190	<6	no	-.23 (4)	O	-.23 (4)	-	74	≥6	no	-.42 (4)	-	-
11	Barker	1998	-	-	-	-	-	-	-	-	-	-	-	-	-
12	Tellier	1998	-	-	-	-	-	-	-	72	<6	yes	.24 (3)	-	-
13	Pullen	1996	31	<6	yes	-.14 (12)	S	-.14 (12)	-	-	-	-	-	-	-
14	Leeman et al.	1993	39	<6	yes	.70 (4)	S	.70 (4)	-	-	-	-	-	-	-
15	Elrod & Minor	1992	-	-	-	-	-	-	-	-	-	-	-	-	-
16	Garrido & Sanchis	1991	-	-	-	-	-	-	-	20	<6	yes	-1.13 (6)	-	-
17	Steele	1991	-	-	-	-	-	-	-	-	-	-	-	-	-
18	Guerra & Slaby	1990	80	<6	no	1.06 (24)	S	-	1.06 (24)	-	-	-	-	-	-
19	Shivratna	1988	29	<6	yes	.30 (112)	S (56), O (56)	.31 (96)	.23 (16)	29	<6	yes	-.04 (8)	-	-
20	Botcher	1985	-	-	-	-	-	-	-	-	-	-	-	-	-
21	Long & Sherer	1985	18	<6	yes	.52 (6)	S (4), O (2)	.31 (4)	.95 (2)	-	-	-	-	-	-
22	Bowman & Auerbach	1982	10	<6	yes	.51 (5)	O	-	.51 (5)	-	-	-	-	-	-
23	Schlichter & Horan	1981	19	<6	yes	.41 (2)	O	-.52 (2)	-	-	-	-	-	-	-
24	Toftte-Tippis	1980	20	<6	yes (12), no (6)	.33 (18)	S	.40 (12)	.20 (6)	-	-	-	-	-	-
25	Feindler	1979	8	<6	yes	.54 (8)	S (1), O (7)	-.05 (5)	1.51 (3)	-	-	-	-	-	-
26	Green-Burns	1979	10	<6	no	.00 (10)	O	-.01 (5)	.01 (5)	-	-	-	-	-	-
27	Ollendick & Hersen	1979	18	<6	no	.76 (28)	S (2), O (26)	.72 (26)	1.33 (2)	18	<6	no	1.52 (2)	-	-
28	Sarason & Ganzer	1973	128	<6	no	.15 (2)	O	.15 (2)	-	-	-	-	-	-	-

Note. n = sample size; FU (#ES) = follow-up duration in months (number of effect sizes); pre = controlled for pretest score yes/no; pd (#ES) = mean effect size (number of effect sizes); inf (#ES) = informant is self (S) or others (O) (number of effect sizes); Construct d (#ES) = average effect size (number of effect sizes) measuring prosocial behavior (PRO), and problem-solving skills (PSS).