How effective is the "Reasoning and Rehabilitation" (R&R) program in changing cognitive and behavioral skills? A systematic review and meta-analysis[☆]

Olga Sánchez de Ribera [a,*], Violeta Chitgian Urzúa [b], Genée Pienaar [a,c]

[a] *Department of Criminology, The University of Manchester, United Kingdom*
[b] *Department of Criminology, University of Ottawa, Canada*
[c] *HM Prison and Probation Service, United Kingdom*

ABSTRACT

Interventions for individuals who commit offenses are of great importance to reduce criminal recidivism by targeting criminogenic factors. The first and most widely applied program is the Reasoning & Rehabilitation (R&R) program. Despite evidence that the R&R program (and its derivatives) is effective in reducing recidivism, questions remain regarding the benefits in a range of cognitive and behavioral outcomes, the long-term effects, and the difference between psychosocial outcomes for different individuals' characteristics (i.e., sex, age, mental disorders, intellectual disabilities). This systematic review and meta-analysis address these issues. A total of 28 studies were eligible for inclusion in the systematic review and 23 studies ($N = 2528$) for the meta-analysis. Results indicated that the R&R is effective in increasing (social) problem solving ($SMD = 0.26$, $p = 0.009$), empathy/social-perspective taking ($SMD = 0.37$, $p < 0.001$), and decreasing violence/aggression ($SMD = 0.38$, $p = 0.003$), anger/hostility ($SMD = 0.25$, $p = 0.003$), and impulsivity/inhibition ($SMD = 0.27$, $p = 0.003$) but not on criminal attitudes ($SMD = 0.20$, $p = 0.07$). Secondary, and some other primary outcomes, were not examined owing to the small number of studies that included these outcomes. We conclude that the R&R is effective at improving some psychosocial skills among individuals who commit offenses. However, questions still remain (i.e., the long-term effect on some outcomes, the effect on different types of offenders, and different comparison groups) because of the small number of studies.

1. Introduction

Recidivism is a worldwide issue, with rates ranging from 20% to 63% (Yukhnenko et al., 2019), generating substantial costs to societies (e.g., Aos et al., 2001; Heeks et al., 2018; Jaitman, 2019). Among the different approaches to decrease the rates of recidivism (and their costs), the rehabilitation of individuals who commit offenses is one involving the implementation of cognitive-behavioral programs. The risk-needs-responsivity model (RNR; Bonta & Andrews, 2017), which underlies the rehabilitation approach of the individuals who commit offenses, postulates that the program intensity should be tailored to the level of risk of the treated individuals (*risk principle*), programs should target the

dynamic risk factors associated with recidivism (*needs principle*), and the program should be tailored to the characteristics of the individuals who commit offenses such as age, sex, and level of motivation to increase its effectiveness (*responsivity principle*; Bonta & Andrews, 2017). A prototypical example is the Reasoning and Rehabilitation (R&R) program (Ross & Fabiano, 1985), which aims to reduce reoffending by changing individuals' thinking style and attitudes through developing individuals' prosocial skills. However, studies on the effectiveness of this program have mainly focused on recidivism as an outcome (for meta-analysis, see Tong & Farrington, 2006), so the effect of the program on prosocial behaviors, thinking and attitudes needs further examination. Examining the effect of the program on therapeutic outcomes

would help explain how these outcomes are related to the final outcome, recidivism.

1.1. The Reasoning and Rehabilitation (R&R) program

Based on social learning theory (Bandura, 1977), cognitive behavioral therapy (CBT) asserts that individuals' antisocial values, beliefs and behaviors are learned through observation and subject to vicarious reinforcement in their environment (i.e., family, peers, school, media). A considerable number of studies have recognized the influence of cognitive malfunctioning on the onset and maintenance of antisocial behavior (Walters, 2022). Thus, programs for individuals who commit offenses are aimed at converting antisocial thinking and attitudes into prosocial thinking by targeting the development of skills in problem-solving, social skills training, and prosocial modeling with the positive reinforcement of non-criminal behaviors and attitudes. Evidence shows that cognitive behavioral approaches are an effective method of reducing recidivism among different types of individuals who commit offenses (e.g., Lipsey et al., 2007; Pearson et al., 2002; Redondo et al., 1999; Schmucker & Lösel, 2017; Wilson et al., 2005).

A prototypical example of this type of program is the Reasoning and Rehabilitation (R&R) program, which was developed by Ross & Fabiano in the 1980s in Canada (for a detailed description, see Ross & Fabiano, 1985) and is now widely implemented in criminal justice systems worldwide to reduce the risk of reoffending (Antonowicz, 2008). The R&R is a manual-driven and multi-faceted program in which participants learn skills such as impulse control, critical reasoning, rational thinking, socially acceptable interactions, perspective taking, self-control and prosocial problem-solving (Ross & Hilborn, 2008). Although there are different versions of the R&R, depending on the target group of individuals who commit offenses (i.e., youths, adults, women, drug users, inmates with Attention Deficit and Hyperactivity Disorder (ADHD) and the mentally disordered), the R&R usually consists of 36 2-hour group sessions that employ various techniques such as role-play, modeling, discussion, individual exercises and practice in real situations to consolidate the new skills (Robinson & Porporino, 2001; Ross et al., 1988). The latest version to date, R&R2, is reduced to 14 sessions and incorporates specific techniques for fostering prosocial neurodevelopment based on the neurocriminology model (Ross & Hilborn, 2008).

1.2. Existing reviews of the R&R program

Over the 40 years of R&R implementation in 17 countries, several reviews have been published addressing the effectiveness of the R&R in reducing recidivism (for detail, see Tong & Farrington, 2006), but only one is a meta-analysis (Tong & Farrington, 2006). This meta-analysis, based on four countries (Canada, USA, UK and Sweden), showed a significant 14 % decrease in recidivism for program participants compared with controls. However, it remains unknown whether the R&R has an impact on the criminogenic needs targeted (e.g., antisocial attitudes, impulsivity, perspective taking). Although several studies have assessed the impact of R&R on treatment outcomes using different validated measures, the results are mixed (see Antonowicz, 2005) and a meta-analysis has not yet been conducted. The results of a meta-analysis may have significant impacts on practice, policy, and research. Psychologists in criminal justice agencies, managers and policymakers can use the findings to improve the performance of R&R programs by addressing gaps within this program. Second, it may provide a foundation for developing a standardized outcome measure of the program.

1.3. The current study

The current review aims to provide comprehensive and systematic evidence of the efficacy of the R&R and its derivatives in altering different psychosocial outcomes targeted by the program. Furthermore,

the question of which factors, such as the characteristics of participants, settings, and methodology, have an impact on the effectiveness of R&R will also be addressed. Finally, the long-term efficacy of the R&R on the outcomes examined across the studies will be examine.

2. Methods

This review adhered to the Preferred Reporting Items for Systematic Review and Meta-Analysis (PRISMA, Moher et al., 2009). A protocol was developed but not registered.

2.1. Eligibility criteria

To be included in this meta-analysis, studies must have met the following inclusion criteria:

1. *Participants*: studies with data from all age groups, any sex, any nationality, any type of individual who committed offenses (including those with intellectual disabilities, ADHD, and mental disorders) and in any setting (i.e., forensic institution, prison, probation, community) were included. Samples from the general population were excluded.
2. *Intervention*: only the R&R program as well as any of its versions [e.g., R&R2, R&R2MHP, R&RADHD] and derivatives [e.g., Thinking Skills Program (TSP), Enhanced Thinking Skills Program (ETSP)] were included. Cognitive behavioral programs were excluded if they did not explicitly refer to R&R (or one of the derivatives) or the creators of the program (i.e., Ross & Fabiano).
3. *Study design*: the studies included used either a randomized or quasi-experimental design that compared the treated group (i.e., R&R) with a control group. Both inactive control groups (e.g., no treatment, waitlist) and active control groups (e.g., other treatment, treatment as usual) were considered as control conditions. Studies were excluded if the control group was a derivative program of the R&R (e.g., ETSP, TSP), a group of non-completers (versus completers), and those without a control group.
4. *Outcome*: studies had to report sufficient data for the calculation of the effect sizes [e.g., *Mean* (SD)] to be included in the meta-analysis, and studies had to provide information on at least one of the following outcomes:
   4.1 *Primary outcome*: the primary outcome was any of the cognitive-behavioral skills targeted by the program including violence, prison misconduct, impulsivity, social and interpersonal skills, critical and creative thinking, antisocial values, assertiveness, negotiation skills, social perspective taking, and level of risk of reoffending. These outcomes are assessed with a validated instrument. Studies that reported only recidivism rates were excluded (see Tong & Farrington, 2006).
   4.2 *Secondary outcomes*: the secondary outcome measures were quality of life such as achieving/maintaining a job, decreasing drug and alcohol consumption, managing financial debts, well-being, depression, and anxiety.
      Where possible the outcomes were classified as short-term (up to 6 months), medium-term (from 7 to 12 months) and long-term (>12 months).
5. *Language*: only studies in English and Spanish were included.
6. *Publication type*: peer-reviewed articles, government reports and PhD dissertations were included in both published and unpublished form, but books and book chapters were excluded.

2.2. Search strategy

In order to identify relevant studies, three electronic databases were searched: PubMed, Scopus and Web of Science up to 2021. The initial search was conducted from February to May 2021, and updated in January 2024. The search for candidate studies to be included in the

meta-analysis was conducted using keywords describing participants (e. g., "*offenders*" OR "*prison*" OR "*mentally disordered*"), intervention (e.g., "*Rehabilitation and Reasoning*" OR "*cognitive program*"), design (e.g., "*randomized control* trial*" OR "*trial*" OR "*quasi-experimental*"), outcome (e.g., "*criminal attitudes*" OR "*impulsivity*" OR "*cognitive abilities*" OR "*social skills*"), and Medical Subject Headings (MeSH). The search terms were truncated, used separately and in different combinations for the database searches, and were searched by title, abstract or keyword (see Appendix 1 for the specific search syntax).

In addition, the reference lists of the included studies and relevant systematic reviews were revised to identify any additional studies missed from the electronic searches. Finally, one of the program developers, Prof. Robert Ross, was contacted for information on (un) published or ongoing studies, and the main authors were contacted by email to request data or additional information about the potential primary studies that could be included in this systematic review and meta-analysis.

2.3. Study selection

Applying the inclusion and exclusion criteria, two researchers screened titles and abstracts independently, after duplicate removal. Two review authors (OSR, VCU) independently screened the full texts to identify studies for inclusion, and recorded reasons for exclusion of the ineligible studies. Any disagreements were resolved through discussion.

2.4. Data management and extraction

Search results from electronic databases and manual searches were exported to Microsoft Excel. Data was independently extracted by two reviewers (OSR, VCU) using a standardized extraction form in Microsoft Excel which included bibliographical information, sample and program characteristics, methodological features, study quality, outcomes and time points recorded, and data for the calculation of effect sizes. The first author developed the coding protocol, the second author coded three studies to establish consistent coding guidelines. Thereafter, both authors independently coded each of the studies and coding disagreements were solved through discussion between both authors.

2.5. Risk of bias assessment in the studies

Two reviewers (GP, OSR) independently assessed the risk of bias for each study (i.e., RCT and non-RCT) using the EPOC tool (EPOC, 2014). This tool was selected to assess non-RCT because it is most suitable for non-medical trials and is based on the Campbell Collaboration guidelines (Gaffney et al., 2021). The risk of bias were assessed in the following domains: allocation sequence, allocation concealment, baseline equivalence, baseline characteristics, incomplete outcome data, blind outcome assessment, contamination protection and selective outcome reporting (for an explanation of each domain, see Table 5 in Gaffney et al., 2021). Studies were rated as 'low risk', 'high risk' or 'unclear risk' for each domain, but, in this study, the overall score was removed following Higgins' et al. (2011) recommendations.

Before starting the assessments, the questions and domains were thoroughly discussed so that both reviewers had a common understanding and consistent coding guidelines. Cohen's kappa statistic (Cohen, 1960) was used to assess the level of agreement between two raters as it considers chance levels of agreement and is appropriate for categorical items. The kappa score ranges from −1 to 1, where 1 indicates a perfect level of agreement between the two raters. Disagreements were resolved through discussion. The studies for meta-analysis were not excluded based on the 'risk of bias' assessment. However, sensitivity analyses for the outcomes were conducted, which excluded trials with high or unclear risk of bias ratings for allocation concealment, or for blinding of outcome assessment.

2.6. Statistical analysis

Meta-analyses were only conducted where this was meaningful, that is, if there were at least five studies per outcome (Jackson & Turner, 2017) (For the individual effect size for each outcome per study, see Appendix 2). Effect sizes (ESs) were computed and pooled the individual ESs with Cochrane Collaboration Review Manager software (*RevMan 5.1; Review Manager*, 2014). For each outcome, the standardized mean difference (SMD or *Hedges' g*) at post-test and follow-up were calculated by subtracting the mean score of the comparison group from the mean score of the R&R program group and dividing the result by the pooled and weighted standard deviation of the two groups (Borenstein et al., 2009). When higher scores in an instrument indicated a greater degree of problems (e.g., higher levels of anger or violence), the mean was multiplied by −1. Thus, a positive effect size indicated treatment effects favoring the group receiving the R&R program. *SMD* values of 0.2, 0.5, and 0.8 were considered thresholds for small, moderate, and large effect sizes respectively (Cohen, 1988). In studies missing the statistical information to calculate the effect size, the authors were contacted to obtain such information. When that was not possible, the study was excluded from the meta-analysis but included in the systematic review.

Heterogeneity tests such as Chi^2 (*Q* static; Deeks et al., 2023b) and I^2 (Higgins et al., 2003) were used to determine the variation in outcomes between studies. The latter describes the percentage of variation across studies that are due to heterogeneity rather than chance. Unlike *Q*, it does not inherently depend upon the number of studies considered (Higgins et al., 2003). The random-effect model was applied to analyse data across the studies. Subgroup and sensitivity analyses were performed when heterogeneity was significantly moderate (I^2 = 30 % to 60 %) or high (I^2 = 60 % to 90 %) (Higgins et al., 2003).

Publication bias for outcomes with at least 10 studies was performed (Page et al., 2023), using funnel plots under the fixed effect model (Sterne et al., 2017).

2.7. Subgroup analyses

In the unregistered protocol, we planned to undertake subgroup analyses by sex (male, female and mixed) and age (adults, adolescents/ children) but found that there was not much variability across the studies. Consequently, we did not consider these two variables. When it was possible and meaningful to do so, subgroup analyses were undertaken to examine differences between settings (prison, forensic unit, other), design (RCT, non-RCT), and whether the authors affiliated with the program conducted the evaluation (yes, no). To ensure subgroup analyses were meaningful, these analyses were undertaken where there was data from at least three studies for two main subgroups (James et al., 2020). For these categorical variables, a series of the analogue to ANOVA (with random effect model) were conducted. Additionally, we used an alternative method for testing for differences between subgroups, meta-regression. Jamovi version 2.2 software (The Jamovi Project, 2021) was used to examine whether categorical moderators such as the study design, setting, and authors' affiliation to the program had an impact on the effect of the program on the psychosocial outcomes. Meta-regression analyses were computed only if 10 studies were included in the meta-analysis (Deeks et al., 2023a), without outliers under the random-effect model and using the restricted maximum likelihood (REML) method (Tanriver-Ayder et al., 2021).

3. Results

3.1. Study selection

The database search identified 2830 records. An additional 114 records were identified from searching reference lists and correspondence with the program developer (Prof. Ross, and the reference lists). After removing duplicates, researchers screened 2415 titles/abstracts, and

excluded 2166 records. Following our inclusion criteria, we assessed 250 full texts for eligibility for inclusion in the review and 29 eligible articles were identified. Two of these articles (Young et al., 2017, 2015) used the same sample, we thus treated them as one study (i.e., 28 studies and 29 articles). Five of these articles were excluded from the meta-analysis (but not from the systematic review) because the outcomes were included in less than five studies, the authors did not have the data to calculate the effect size, or did not respond to our email communication. Therefore, 28 studies ($n = 4544$) met the criteria for inclusion in the systematic review, and 23 studies ($n = 2528$) were included in meta-analyses comprising a total of 1210 R&R participants and 1318 controls. The study selection process is summarized in Fig. 1 below.

3.2. Characteristics of included studies

3.2.1. Participant characteristics

Included studies comprised mainly adults ($k = 19$, 68 %) and male ($k = 22$, 78 %); one study included women (3 %), four studies (14 %) included adolescents, and three studies (11 %) included mixed populations. These participants were mainly from the UK ($k = 12$, 43 %) and from Europe [Germany, Spain ($k = 3$ each, 11 %), Iceland, Netherlands, Switzerland, Sweden ($k = 1$ each, 3 %)], Canada, Chile, Iran ($k = 1$ each, 3 %) and the USA ($k = 3$, 11 %). Most evaluations were conducted in forensic security units ($k = 12$, 43 %) and prison ($k = 11$, 39 %) settings, but two studies involved participants on probation (7 %) and Board of Pardons and Parolees (7 %), and one study in an education and care

Fig. 1. PRISMA flow diagram of the search strategy.

centre (3 %).

3.2.2. Methodology characteristics

Regarding the study methodology, eleven studies (39 %) were RCT's and seventeen (61 %) were quasi-experimental studies. The statistical power was calculated in ten studies (36 %; Cornet et al., 2016; Cullen et al., 2012b; Cullen et al., 2012a; Kingston et al., 2018; Jotangia et al., 2015; McDougall et al., 2009; Rees-Jones et al., 2012; Droppelmann et al., 2020; Yip et al., 2013; Young et al., 2017, 2015). The calculated sample size varies across these studies, but the most common sample size was 35 participants per group. However, six studies (60 %) did not achieve the target sample size (Cornet et al., 2016; Cullen et al., 2012b; Kingston et al., 2018; Jotangia et al., 2015; Sánchez de Ribera, 2015; Young et al., 2017, 2015), that is, the power was too low to determine effectiveness. Four studies (Baggio et al., 2020; Doyle et al., 2013; Redondo et al., 2012; Droppelmann et al., 2020) did not report the power calculation, although three studies (Baggio et al., 2020; Redondo et al., 2012; Sánchez de Ribera, 2015) acknowledged a small sample size and another study acknowledged adequate power (Doyle et al., 2013). Furthermore, seven out of twenty-eight studies (25 %) included a total sample size of <35 participants (range $n = 16$–1347).

3.2.3. Program characteristics

Twelve studies (43 %) applied the R&R and R&R2 versions, whereas six studies (21 %) applied versions of the program for individuals with ADHD and mental disorders. Conversely, 10 studies (36 %) applied derivatives of the R&R. Although the fidelity of the program was not exhaustively assessed in most of the studies, the fidelity was deemed acceptable in 13 studies (46 %), whereas one study reported several issues in the fidelity of the program. Half of the studies ($k = 14$) did not provide information on this topic. The program attrition rate ranged from 11 % to 52 % (average: 25 %) and 7 studies (25 %) did not report the attrition rate. Ten studies (36 %) included follow up periods, whereas more than half of studies did not ($k = 18$, 64 %). In addition, ten studies (36 %) reported that author(s) affiliated to the program were involved in the evaluation of the program (see Table 1). Finally, most of the studies reported mixed results, that is, improvements in some skills but not in others.

3.3. Risk of bias

A summary of the risk of bias for each study is provided in Appendix 2. A significant agreement was found between raters (Cohen's k was 0.59, $p < 0.001$). Some items (SOR) achieved a $k = 1$. Fig. 2 displays the results of the risk of bias analysis for each of the items on the EPOC tool. The description of each risk of bias item is described below.

3.3.1. Allocation sequence generation

We considered seven studies (25 %) as high risk on how participants were assigned to the groups, whereas eight studies (29 %) were determined to be low risk and in 13 studies (46 %) the level of risk was unclear.

3.3.2. Allocation concealment

We categorized four studies (14 %) and 18 studies (64 %) as high and unclear risk, respectively. A further six studies (21 %) were low risk for allocation concealment.

3.3.3. Baseline equivalence in the outcome measure

A high risk of bias was assigned to 10 studies (36 %) for differences in any outcome at baseline, whereas 14 studies (50 %) and four studies (14 %) were low and unclear risk, respectively.

3.3.4. Baseline equivalence in other participant characteristics

Overall, we classified 13 studies (46 %) as low risk on the baseline equivalence in participants' characteristics, whereas seven studies (25
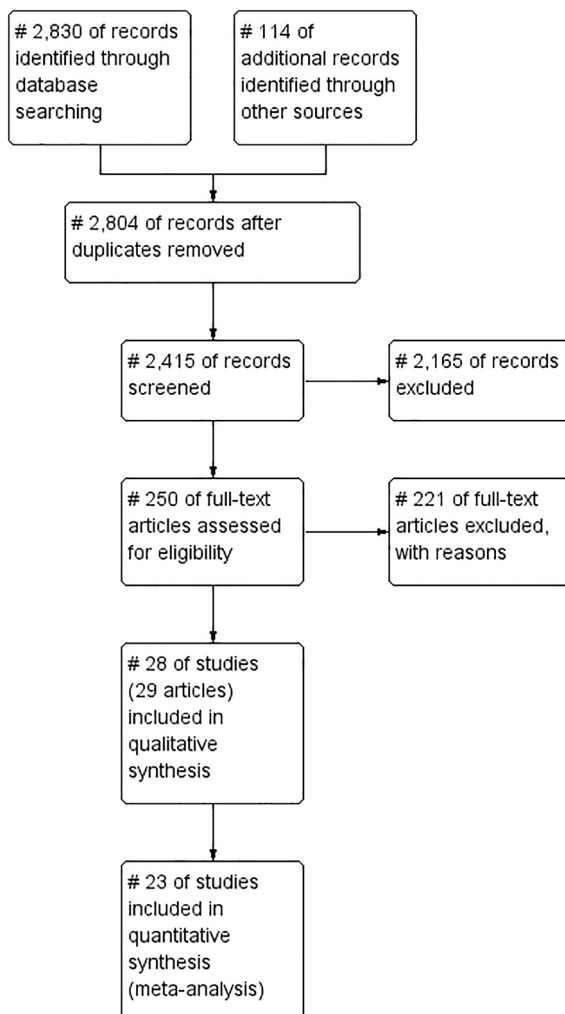
Table 1

Summary of characteristics of included studies in the systematic review and the meta-analysis.

Author(s) (year)	Country setting	Author(s) affiliated to the program involved in evaluation	Integrity of treatment fidelity	Study design	Statistical power calculated	Treatment, Control group type, (*n*)	Sample (age, sex)	Program name	Number of sessions	Program dropouts (%)	Follow up (months)
Baggio et al., 2020*	Switzerland Prison	No	NR	QE	No	T:129 C(TAU):84	Adults Male	R&R2	14	NR	None
Berman, 2004*	Sweden Prison	No	NR	QE	No	T:212 C(NR):451	Adults Male	R&R	NR	23%	None
Clarke et al., 2010*	UK Forensic Security Unit	No	Acceptable	QE	No	T:18 C(TAU):17	Adults Male	R&R	NR	17%	None
Cornet et al., 2016*	Netherlands Prison	No	NR	QE	Yes	T:74 C(WL):53	Adults Male	CoVa/ETS	20	30%	None
Cullen et al., 2012a*	UK Forensic Security Unit	No	Acceptable	RCT	Yes	T:44 C(TAU):40	Male Adults	R&R	36	50%	12
Cullen et al., 2012b*	UK Forensic Security Unit	No	Acceptable	RCT	Yes	T:42 C(TAU):40	Male Adults	R&R	36	52%	12
Curran & Bull, 2009*	UK Education and Care Centre	No	NR	RCT	No	T:14 C(WL):14	Adolescents Boys	Ross Programme	NR	14%	None
Droppelmann et al., 2020*	Chile Prison	Yes	Acceptable	RCT	No	T:26 C(TAU):41	Adult Male	R&R2	14	25%	None
Doyle et al., 2013*	UK Prison	No	NR	QE	No	T:70 C(TAU):56	Male Adults	ETS (Enhanced Thinking Skills)	20	NR	None
Garrido & Sanchis, 1991	Spain Prison	Yes	NR	QE	No	T:14 C:13	NR Adolescents	Psychosocial competence program	12	50%	3
Garrido Genovés & Piñana, 1996*	Spain Prison	Yes	NR	QE	No	T:9 C:18	Male Adults	Prosocial Thinking	38	NR	12 18
Gretenkord, 2004*	Germany Forensic Security Unit	NR	NR	QE	No	T:11 C:5	Male NR	R&R	NR	NR	None
Jotangia et al., 2015*	UK Forensic Security Unit	Yes	Acceptable	QE	Yes	T:18 C(TAU):20	Female Adults	R&R2MHP	36	11%	3
Khodayarifard et al., 2010	Iran Prison	No	NR	RCT	No	T1:48 T2:46 C(WL):40	Adults Male	R&R	16	NR	None
Kingston et al., 2018*	Canada Prison	No	NR	RCT	Yes	T:44 C(TAU):36	Adults Male	R&R2	14	16%	None
McDougall et al., 2009*	UK Prison	No	NR	RCT	Yes	T:229 C(WL):179	Adult Male	ETS	20	4%	3
Pullen, 1996	US Probation	Yes	Several shortfalls reported	RCT	No	T:20 C:20	Adolescents Male	R&R	36	NR	None
Redondo et al., 2012*	Spain Probation	No	NR	QE	No	T:17 C(WL):11	Adolescents Mixed (F:10 M:23)	PTP	40 hours	15%	None
Rees-Jones et al., 2012*	UK	Yes	Acceptable	QE	Yes	T:52 C(TAU):45	Adults Male	R&R2MHP	16	22%	3

Table 1 (*continued*)

Author(s) (year)	Country setting	Author(s) affiliated to the program involved in evaluation	Integrity of treatment fidelity	Study design	Statistical power calculated	Treatment, Control group type, (*n*)	Sample (age, sex)	Program name	Number of sessions	Program dropouts (%)	Follow up (months)
Sánchez de Ribera, 2015*	Forensic Security Unit UK	No	NR	QE	Yes	T:29 C (WL):39	Adult Male	TSP	19	NR	None
Van Voorhis et al., 2001a	Prison USA Board of Pardons and Parolees	No	Acceptable	RCT	No	T: 232 C(TAU): 236	Adults Male	Georgia Cognitive Skills Program	35	40%	3, 6, 9
Van Voorhis et al., 2001b	USA Board of Pardons and Parolees	No	Acceptable	RCT	No	Parolees: T:574 C (TAU):581 Pre-release: T: 104 C(TAU):88	Adults Male Mixed (Male:49% Female:5%)	Georgia Cognitive Skills Program	35	31%	3, 6, 9, 12
Wettermann et al., 2012*	Germany Forensic Psychiatric Clinic	No	NR	QE	No	T:14 C(TAU):17	Adult Male	R&R	36	20%	None
Wettermann et al., 2020*	Germany Forensic Psychiatric Hospital	No	Acceptable	QE	No	T:47 C(TAU):28 C(DBT-F):34	Adult Male	R&R	36	11%	None
Yip et al., 2013*	UK Forensic Security Unit	Yes	Acceptable	QE	Yes	T:30 C(TAU):29	Adult Male	R&R2MHP	16	20%	None
Young et al., 2010*	UK Forensic Security Unit	Yes	Acceptable	QE	No	T:34 C(WL):12	Adult Male	R&R2M	16	35%	None
Young et al., 2012*	UK Forensic Security Unit	Yes	Acceptable	QE	No	T:16 C(TAU):15	Adult Male	R&R2 ADHD	15	24%	None
Young et al., 2015, 2017*	Iceland Forensic Security Unit	Yes	Acceptable	RCT	Yes	T:33 C(TAU):34	Adult Mixed (F:62 M:33)	R&R2 ADHD	15	48%	3

Note: ADHD: Attention deficit and hyperactivity disorder; DBT-F: Dialectical Behavioral Therapy– Forensic; EF: Executive Functioning; ETS: Enhanced Thinking Skill program; F: Female; M: Male; NR: Not reported; PTP: Prosocial Thinking Program; QE: Quasi-experimental; RCT: Randomized controlled trial; R&R: Reasoning & Rehabilitation; TAU: Treated as usual; TSP: Thinking Skills Program; UK: United Kingdom; US: United States; WL: Wait list.

%) and eight studies (29 %) were categorized as high and unclear risk, respectively.

3.3.5. Incomplete outcome data

Only one study (4 %) was categorized as high risk in differential attrition. Most of the studies, 14 studies (50 %) and 13 studies (46 %) were judged to be low risk and unclear, respectively.

3.3.6. Blinding of outcome assessment

Overall, we classified 18 studies (64 %) as low risk, whereas 8 studies (29 %) and two studies (7 %) were unclear and high risk, respectively, on how individuals collected the data.

3.3.7. Protection against contamination

Several studies reported that experimental groups were in the same setting but measures to avoid contamination were not mentioned, so this

item was categorized as "unclear" in level of risk, resulting in 26 studies (93 %) being categorized as unclear. One study (4 %) was categorized as high and another study as low risk (4 %).

3.3.8. Selective outcome reporting

The assessment of this item was difficult because we did not have access to the protocols of the studies included in this review. Consequently, we classified the majority of trials as having an unclear risk of bias for this item (24 studies; 86 %). However, we found a high risk of SOR in two studies (7 %) even though we could not access the protocol, and another two studies were judged as low risk (7 %).

3.4. Effect of the program: meta-analysis

After accounting for missing information, and outcomes included in less than five studies (i.e., disruptive behavior, coping strategies, locus of
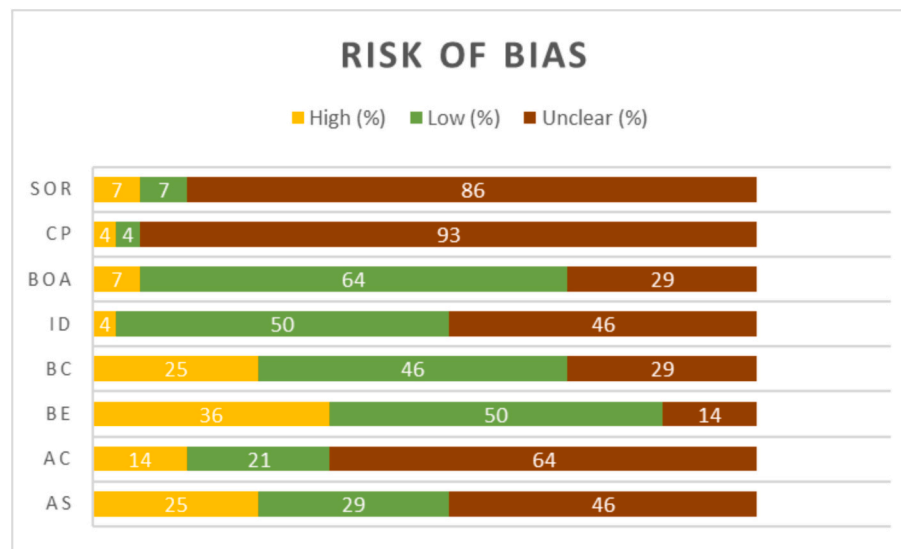
Fig. 2. Risk of bias in included studies based on EPOC risk of bias tool. *Note*: AC, allocation concealment; AS, allocation sequence; BC, baseline equivalence on participant characteristics; BE, baseline equivalence on outcomes; BOA, blind outcome assessment; CP, contamination protection; ID, incomplete outcome data; SOR, selected outcome reporting.

control, ADHD, self-esteem, anxiety, depression, life quality and satis-faction, interpersonal problems, accepting responsibility, substance use in prison, external attribution, sense of coherence and executive func-tioning), a total of 23 studies were eligible for inclusion in our meta-analysis. The overall effects of the R&R found for each primary outcome are described below (see Table 2). Subgroup analyses were only conducted if there was evidence of significant heterogeneity and if there were enough studies.

Regarding the short, medium and long-term effects, when we examined the impact of the R&R (i.e., >3 months after finishing the program), the benefits of the interventions were less clear. The effect was non-statistically significant for violent behavior ($SMD = -0.14$; 95 % CI: -0.34 to 0.07) and anger/hostility ($SMD = 0.11$; 95 % CI: -0.19 to 0.41). However, these findings must be interpreted with caution since the analysis is based on three and two studies respectively, and the follow up period was from 3 to 12 months.

Finally, there were not enough studies or data available for exam-ining the effect of the program on secondary outcomes included in our protocol such as job attainment, drugs/alcohol consumption and financial management. Three studies reported life quality (Young et al.,

2015), life satisfaction (Young et al., 2017) and substance use in prison (Cullen et al., 2012b) but these outcomes were not included in the meta-analysis because they were reported in less than five studies (for the effect size of these and other outcomes, see Appendix 3). Below we describe the results of the meta-analysis for the primary outcomes most reported across the studies.

3.4.1. Effect of the R&R on (social) problem solving

Fourteen studies (Clarke et al., 2010; Cullen et al., 2012a; Curran & Bull, 2009; Doyle et al., 2013; Garrido Genovés & Piñana, 1996; Gre-tenkord, 2004; Jotangia et al., 2015; McDougall et al., 2009; Redondo et al., 2012; Rees-Jones et al., 2012; Wettermann et al., 2012; Yip et al., 2013; Young et al., 2012, 2010) were included in the analyses examining an improvement in social problem solving post-treatment. A significant small effect for the treatment group was found ($SMD = 0.41$, 95 % CI: 0.14 to 0.68; $p = 0.003$), with evidence of moderate heterogeneity ($\chi^2 = 40.30, p < 0.001, I^2 = 68$ %). When we added the three studies that used executive functioning tasks to measure problem solving skills (Cornet et al., 2016; Sánchez de Ribera, 2015; Wettermann et al., 2020), a sig-nificant small effect for the treated group was found ($SMD = 0.26$, 95 % CI: 0.07 to 0.45; $p = 0.009$), with significant heterogeneity ($\chi^2 = 47.31$, $p < 0.001$, $I^2 = 62$ %). Additionally, when we conducted a sensitivity analysis removing three outliers (Clarke et al., 2010; Curran & Bull, 2009; Garrido Genovés & Piñana, 1996) the effect remained significant ($SMD = 0.12$, 95 % CI: 0.00 to 0.24; $p = 0.04$) with non-significant heterogeneity ($\chi^2 = 16.07$, $p = 0.38$, $I^2 = 7$ %) (See Fig. 3). After excluding trials with high or unclear risk of bias ratings for allocation concealment or for blinding of outcome assessment (Clarke et al., 2010; Cornet et al., 2016; Curran & Bull, 2009; Garrido Genovés & Piñana, 1996; Gretenkord, 2004; Rees-Jones et al., 2012; Wettermann et al., 2012, 2020; Yip et al., 2013; Young et al., 2012, 2010), the small effect remains significant ($SMD = 0.21$, 95 % CI: 0.04 to 0.37; $p = 0.01$) and the heterogeneity is not significant ($\chi^2 = 6.14$, $p = 0.29$, $I^2 = 19$ %).

Subgroup and meta-regression analyses were not conducted because the heterogeneity was not significant.

3.4.2. Effect of the R&R on impulsivity & inhibition

The thirteen studies (Berman, 2004; Cullen et al., 2012a; Doyle et al., 2013; Droppelmann et al., 2020; Gretenkord, 2004; Jotangia et al., 2015; McDougall et al., 2009; Redondo et al., 2012; Rees-Jones et al., 2012; Wettermann et al., 2012; Yip et al., 2013; Young et al., 2012;

Table 2
Meta-analyses results for each psychosocial and behavioral outcome ($k \geq 5$).

Outcome (RM)	k	*SMD* [95 % CI]	*p value*	Chi^2	I^2 (%)
(Social) Problem solving	17	0.26 [0.07, 0.45]	0.009	47.31***	62
Impulsivity/inhibition	16	0.27 [0.09, 0.45]	0.003	73.27***	71
Violence/aggression	11	0.38 [0.13, 0.63]	0.003	34.27***	68
Anger/hostility	7	0.25 [0.04, 0.46]	0.02	9.12**	34
Empathy/social perspective taking	7	0.35 [0.10, 0.60]	0.007	14.63	38
Criminal attitudes	7	0.20 [−0.01, 0.41]	0.07	18.00*	56

Note: RM: random effect model; *k*: number of studies. These results include outliers.

   * $p < 0.05$.
  ** $p < 0.01$.
 *** $p < 0.001$.

Study or Subgroup	Treatment Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Clarke et al. 2010	75.6	12.9	15	59.3	13.4	17	0.0%	1.21 [0.44, 1.97]
Cornet et al., 2016	-5.23	5	83	-5.64	3.88	50	10.0%	0.09 [-0.26, 0.44]
Cornet et al., 2016	74.42	28.51	84	73.77	30.46	63	11.3%	0.02 [-0.30, 0.35]
Cullen et al. 2012a	13.4	2.2	36	13.4	2.3	36	6.0%	0.00 [-0.46, 0.46]
Curran & Bull, 2009	-90.71	12.16	14	-109.07	15.53	14	0.0%	1.28 [0.45, 2.10]
Doyle et al. 2013	111.72	16.64	70	103.01	12.69	56	9.6%	0.58 [0.22, 0.94]
Garrido & Piñana, 1996	10	2.24	9	4.89	2.05	18	0.0%	2.35 [1.30, 3.39]
Gretenkord, 2004	3.27	1.05	11	2.2	1.54	5	1.1%	0.84 [-0.27, 1.94]
Jotangia et al. 2015	12.19	2.79	18	11.47	3.24	20	3.3%	0.23 [-0.41, 0.87]
McDougall et al., 2009	51.55	6.1	165	50.55	5.85	95	17.5%	0.17 [-0.09, 0.42]
Redondo et al. 2012	93.12	14.19	17	94.82	15.08	11	2.3%	-0.11 [-0.87, 0.65]
Rees-Jones et al. 2012	12.55	2.9	67	12.84	2.46	54	9.6%	-0.11 [-0.46, 0.25]
Sánchez de Ribera, 2015	2.23	1.13	30	2.19	1.15	39	5.7%	0.03 [-0.44, 0.51]
Wettermann et al. 2012	3.22	0.86	14	3.29	1.07	17	2.7%	-0.07 [-0.78, 0.64]
Wettermann et al. 2020	53.41	31.86	46	56.48	31.14	33	6.4%	-0.10 [-0.54, 0.35]
Wettermann et al. 2020	53.41	31.86	46	57.76	27.95	21	4.9%	-0.14 [-0.66, 0.38]
Yip et al. 2013	12.95	3.32	30	11.01	3.84	29	4.8%	0.53 [0.01, 1.05]
Young et al. 2010	13.17	2.8	18	13.47	2.5	9	2.1%	-0.11 [-0.91, 0.69]
Young et al. 2012	13.74	3.34	16	12.35	4.04	15	2.6%	0.37 [-0.34, 1.08]
Total (95% CI)			**751**			**553**	**100.0%**	**0.12 [0.00, 0.24]**

Heterogeneity: Tau² = 0.00; Chi² = 16.07, df = 15 (P = 0.38); I² = 7%
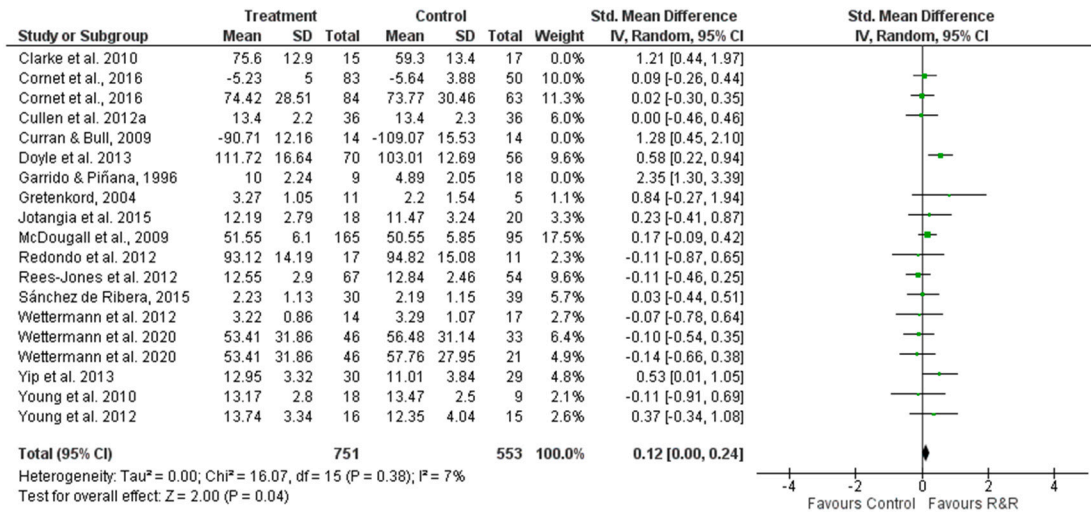Test for overall effect: Z = 2.00 (P = 0.04)

Fig. 3. Forest plot for (social) problem solving outcome. *Note*: Outliers excluded (i.e., Clarke et al., 2010; Curran & Bull, 2009; Garrido Genovés & Piñana, 1996) of this analysis.

Young et al., 2015) included in the analyses showed a reduction in impulsivity post-treatment. We found a significant small effect for the treatment group ($SMD = 0.29$, 95 % CI: 0.07 to 0.51; $p = 0.009$) with significant heterogeneity ($\chi^2 = 61.65$, $p < 0.001$, I² = 74 %). When we included the three studies that assessed inhibition using executive functioning tasks (Cornet et al., 2016; Sánchez de Ribera, 2015; Wettermann et al., 2020), we found a very similar effect ($SMD = 0.27$, 95 % CI: 0.09 to 0.45; $p = 0.003$) with significant high heterogeneity ($\chi^2 = 73.27$, $p < 0.001$, I² = 71 %) (See Fig. 4). Sensitivity analysis excluding studies with risk of bias for allocation concealment or for blinding of outcome assessment (Cornet et al., 2016; Gretenkord, 2004; Rees-Jones et al., 2011; Wettermann et al., 2012; Yip et al., 2013) yield a significant moderate effect ($SMD = 0.30$, 95 % CI: 0.19 to 0.40; $p < 0.001$) with significant heterogeneity ($\chi^2 = 48.41$, $p < 0.001$, I² = 69 %).

Subgroup analyses found that study design ($\chi^2 = 0.02$, $p = 0.90$, I² = 0 %), setting ($\chi^2 = 2.60$, $p = 0.27$, I² = 23 %) and authors' affiliation to the program ($\chi^2 = 3.59$, $p = 0.17$, I² = 44.3 %) did not show significant differences between each subcategory in each group, and meta-regression analyses corroborated it by showing that none of these moderators had an impact on the program effectiveness and

impulsivity/inhibition (study design: $b = 0.03$, SE = 0.20, Z = 0.16, $p = 0.88$; setting: $b = -0.12$, SE = 0.16, Z = -0.76, $p = 0.45$; authors' affiliation: $b = 0.26$, SE = 0.17, Z = 1.53, $p = 0.13$).

Only two studies evaluated the effect of impulsivity after 3 months (Young et al., 2015) and 12 months (Cullen et al., 2012a), and the effect was significantly positive ($SMD = 0.46$, 95 % CI: 0.02 to 0.89; $p = 0.04$) with non-significant heterogeneity ($\chi^2 = 3.66$, $p = 0.16$, I² = 45 %).

3.4.3. Effect of the R&R on violent attitudes, aggressiveness, and antisocial traits

Eleven studies (Baggio et al., 2020; Cullen et al., 2012b; Doyle et al., 2013; Droppelmann et al., 2020; Jotangia et al., 2015; Redondo et al., 2012; Rees-Jones et al., 2012; Yip et al., 2013; Young et al., 2017, 2012, 2010) were included in the analyses examining a reduction of violent attitudes and aggressiveness post-treatment. A significant small effect for the treatment group was found ($SMD = 0.38$, 95 % CI: 0.13 to 0.63; $p = 0.003$), with evidence of high and significant heterogeneity ($\chi^2 = 34.27$, $p < 0.001$, I² = 68 %). Additionally, when removing the outlier (Redondo et al., 2012) the R&R reduced violence significantly in the treated group ($SMD = 0.30$, 95 % CI: 0.09 to 0.51; $p = 0.005$), with
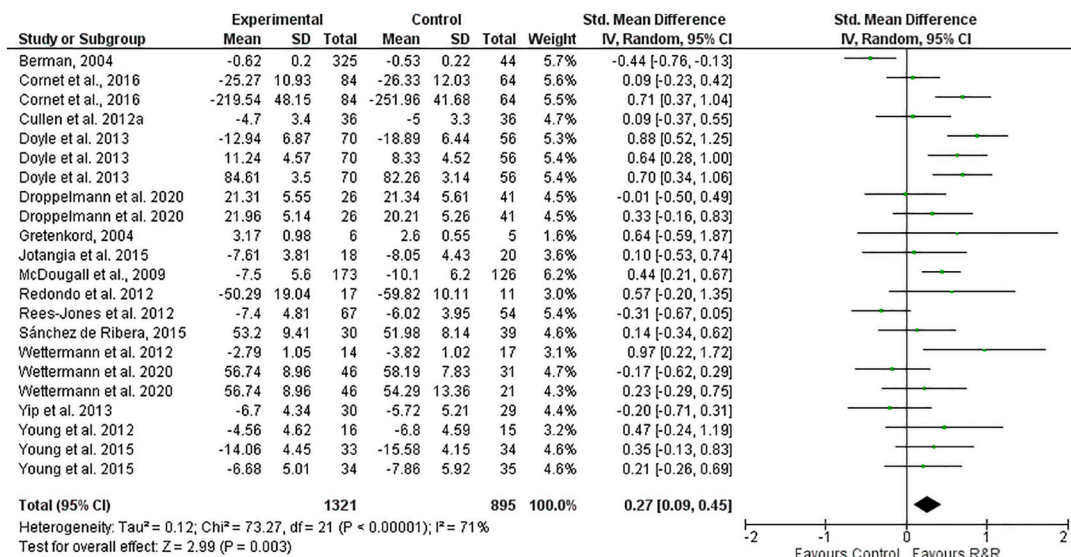
Study or Subgroup	Experimental Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Berman, 2004	-0.62	0.2	325	-0.53	0.22	44	5.7%	-0.44 [-0.76, -0.13]
Cornet et al., 2016	-25.27	10.93	84	-26.33	12.03	64	5.6%	0.09 [-0.23, 0.42]
Cornet et al., 2016	-219.54	48.15	84	-251.96	41.68	64	5.5%	0.71 [0.37, 1.04]
Cullen et al. 2012a	-4.7	3.4	36	-5	3.3	36	4.7%	0.09 [-0.37, 0.55]
Doyle et al. 2013	-12.94	6.87	70	-18.89	6.44	56	5.3%	0.88 [0.52, 1.25]
Doyle et al. 2013	11.24	4.57	70	8.33	4.52	56	5.4%	0.64 [0.28, 1.00]
Doyle et al. 2013	84.61	3.5	70	82.26	3.14	56	5.4%	0.70 [0.34, 1.06]
Droppelmann et al. 2020	21.31	5.55	26	21.34	5.61	41	4.5%	-0.01 [-0.50, 0.49]
Droppelmann et al. 2020	21.96	5.14	26	20.21	5.26	41	4.5%	0.33 [-0.16, 0.83]
Gretenkord, 2004	3.17	0.98	6	2.6	0.55	5	1.6%	0.64 [-0.59, 1.87]
Jotangia et al. 2015	-7.61	3.81	18	-8.05	4.43	20	3.6%	0.10 [-0.53, 0.74]
McDougall et al., 2009	-7.5	5.6	173	-10.1	6.2	126	6.2%	0.44 [0.21, 0.67]
Redondo et al. 2012	-50.29	19.04	17	-59.82	10.11	11	3.0%	0.57 [-0.20, 1.35]
Rees-Jones et al. 2012	-7.4	4.81	67	-6.02	3.95	54	5.4%	-0.31 [-0.67, 0.05]
Sánchez de Ribera, 2015	53.2	9.41	30	51.98	8.14	39	4.6%	0.14 [-0.34, 0.62]
Wettermann et al. 2012	-2.79	1.05	14	-3.82	1.02	17	3.1%	0.97 [0.22, 1.72]
Wettermann et al. 2020	56.74	8.96	46	58.19	7.83	31	4.7%	-0.17 [-0.62, 0.29]
Wettermann et al. 2020	56.74	8.96	46	54.29	13.36	21	4.3%	0.23 [-0.29, 0.75]
Yip et al. 2013	-6.7	4.34	30	-5.72	5.21	29	4.4%	-0.20 [-0.71, 0.31]
Young et al. 2012	-4.56	4.62	16	-6.8	4.59	15	3.2%	0.47 [-0.24, 1.19]
Young et al. 2015	-14.06	4.45	33	-15.58	4.15	34	4.6%	0.35 [-0.13, 0.83]
Young et al. 2015	-6.68	5.01	34	-7.86	5.92	35	4.6%	0.21 [-0.26, 0.69]
Total (95% CI)			**1321**			**895**	**100.0%**	**0.27 [0.09, 0.45]**

Heterogeneity: Tau² = 0.12; Chi² = 73.27, df = 21 (P < 0.00001); I² = 71%
Test for overall effect: Z = 2.99 (P = 0.003)

Fig. 4. Forest plot for impulsivity & inhibition outcome. *Note*: In this analysis, the inhibition tasks and the impulsivity/carelessness subscale of the Social Problem-Solving Inventory and the Antisocial Personality Questionnaire were included.

significant heterogeneity ($\chi^2 = 21.87$, $p = 0.02$, $I^2 = 54$ %) (See Fig. 5), and the studies with risk of bias (Cullen et al., 2012b; Doyle et al., 2013; Jotangia et al., 2015; Redondo et al., 2012; Rees-Jones et al., 2012; Young et al., 2010) showed a significant medium effect size ($SMD = 0.40$, 95 % CI: -0.10 to 0.70; $p = 0.009$) with significantly high heterogeneity ($\chi^2 = 16.66$, $p = 0.01$, $I^2 = 64$ %).

Four studies (Cullen et al., 2012b; Jotangia et al., 2015; Rees-Jones et al., 2012; Young et al., 2017) included a follow up period ranging from 3 to 12 months. In this case the effect size was non-significantly small ($SMD = 0.18$, 95 % CI: -0.09 to 0.44; $p = 0.19$), as well as non-significant in heterogeneity ($\chi^2 = 6.15$, $p = 0.19$, $I^2 = 35$ %).

Subgroup analyses (without the outlier) were conducted to examine the sources of heterogeneity, but none contributed it (setting: $\chi^2 = 0.08$, $p = 0.77$, $I^2 = 0$ %; design: $\chi^2 = 1.54$, $p = 0.21$, $I^2 = 35.2$ %; authors' affiliation to the program: $\chi^2 = 0.01$, $p = 0.10$, $I^2 = 0$ %), and meta-regression analyses corroborated that these variables did not moderate the program effectiveness in reducing violence (study design: $b = 0.34$, SE $= 0.20$, Z $= 1.67$, $p = 0.09$; setting: $b = -0.09$, SE $= 0.24$, Z $= -0.37$, $p = 0.71$; authors' affiliation: $b = 0.02$, SE $= 0.23$, Z $= 0.10$, $p = 0.92$).

3.4.4. Effect of the R&R on anger and hostility

Seven studies (Baggio et al., 2020; Cullen et al., 2012a; Doyle et al., 2013; Jotangia et al., 2015; Rees-Jones et al., 2012; Yip et al., 2013 and Young et al., 2012) were included in the analyses examining a reduction in anger and hostility post-treatment. We found that the R&R program significantly reduced violence and hostility ($SMD = 0.25$, 95 % CI: 0.04 to 0.46; $p = 0.02$) and heterogeneity was small and non-significant ($\chi^2 = 9.12$, $p = 0.17$, $I^2 = 34$ %) (see Fig. 6), therefore subgroup analyses were not conducted. Sensitivity analysis removing studies with risk of bias (Rees-Jones et al., 2012) yielded a significantly higher effect ($SMD = 0.33$, 95 % CI: 0.14 to 0.52; $p < 0.001$) with no significant heterogeneity across the studies ($\chi^2 = 5.94$, $p = 0.31$, $I^2 = 16$ %).

3.4.5. Effect of the R&R on empathy/social perspective taking

Seven studies (Berman, 2004; Cullen et al., 2012a; Droppelmann et al., 2020; Garrido Genovés & Piñana, 1996; Gretenkord, 2004; Redondo et al., 2012; Wettermann et al., 2012) were included in the analyses examining an improvement in empathy and social perspective taking post-treatment. We found that the R&R program significantly improved empathy and social perspective taking ($SMD = 0.35$, 95 % CI: 0.10 to 0.60; $p = 0.007$), and the heterogeneity was not significant across the studies ($\chi^2 = 14.63$, $p = 0.10$, $I^2 = 38$ %) (See Fig. 7). When removing one outlier (Garrido Genovés & Piñana, 1996), the effect remains significant between groups ($SMD = 0.33$, 95 % CI: 0.14 to 0.51; $p < 0.001$), with no substantial heterogeneity among studies ($\chi^2 = 7.70$, $p = 0.46$, $I^2 = 0$ %) (See Fig. 7); subgroup analyses were therefore not undertaken.

Sensitivity analysis removing the studies with risk of bias (Garrido Genovés & Piñana, 1996; Gretenkord, 2004; Wettermann et al., 2012) yielded a similar significant effect ($SMD = 0.37$, 95 % CI: 0.17 to 0.58; p

< 0.001), and the heterogeneity remained non-significant ($\chi^2 = 5.03$, $p = 0.28$, $I^2 = 20$ %).

3.4.6. Effect of the R&R on criminal attitudes

We included seven studies (Berman, 2004; Clarke et al., 2010; Cullen et al., 2012a; Kingston et al., 2018; McDougall et al., 2009, Redondo et al., 2012 and Wettermann et al., 2012) in the analyses examining a reduction in criminal attitudes and sentiments post-treatment. We found a non-significant small effect ($SMD = 0.20$, 95 % CI: -0.01 to 0.41; $p = 0.07$) with significant heterogeneity ($\chi^2 = 18.00$, $p = 0.02$, $I^2 = 56$ %) (See Fig. 8). Sensitivity analysis removing studies with risk of bias (Clarke et al., 2010; Wettermann et al., 2012) yielded a similar effect ($SMD = 0.19$, 95 % CI: -0.03 to 0.41; $p = 0.09$) with significant heterogeneity ($\chi^2 = 14.54$, $p = 0.02$, $I^2 = 59$ %).

Despite the significant heterogeneity, subgroup analysis was only meaningful for the study design, as the authors' affiliation and the setting were not present in enough studies to serve as subcategories. We found no subgroup differences for the study design ($\chi^2 = 0.90$, $p = 0.34$, $I^2 = 0$ %), and the meta-regression analysis was not undertaken because of the small number of studies.

3.5. Publication bias

Publication bias was performed for (social) problem-solving, impulsivity, and violence and aggression. The funnel plots in Fig. 9 indicate that publication bias is present in the analysis of effect sizes of violence and aggressiveness but not for social problem solving and impulsivity/inhibition, as the studies do not fall symmetrically around the mean effect size in the top part of the plot (with low standard errors) (Borenstein et al., 2009).

4. Discussion

To our knowledge, this is the first systematic review and meta-analysis that has examined the effect of the R&R program on the psychosocial skills of individuals who committed offenses. The meta-analysis included 23 studies with a total of 2528 participants and pooled effect estimates suggest small to moderate effect sizes in favor of R&R on a range of outcomes when compared to no treatment, TAU, or other program arms. This review provides evidence that the R&R is partially effective in improving some behaviors and psychosocial skills, as well as in reducing recidivism (Tong and Farrington, 2006). Due to the limited data, it is unclear whether these effects are maintained over time.

Our results comparing the R&R to waiting list/usual/alternative service are consistent with the findings of previous systematic reviews and meta-analyses of the effectiveness of cognitive behavioral programs in reducing reoffending in adolescents, adults, men and women (Galway et al., 2022; Kim et al., 2013; Lipsey et al., 2007; Pearson et al., 2002), which reported small and moderate effects in favor of the treated group.
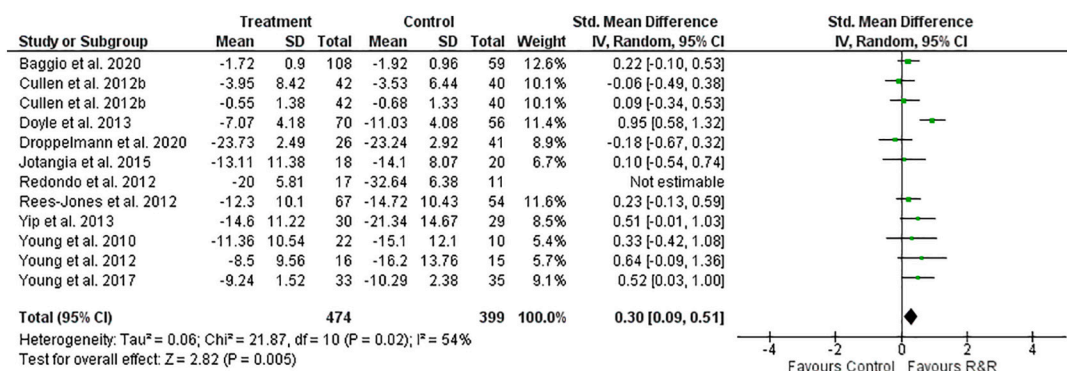
Study or Subgroup	Treatment Mean	SD	Total	Control Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Baggio et al. 2020	-1.72	0.9	108	-1.92	0.96	59	12.6%	0.22 [-0.10, 0.53]
Cullen et al. 2012b	-3.95	8.42	42	-3.53	6.44	40	10.1%	-0.06 [-0.49, 0.38]
Cullen et al. 2012b	-0.55	1.38	42	-0.68	1.33	40	10.1%	0.09 [-0.34, 0.53]
Doyle et al. 2013	-7.07	4.18	70	-11.03	4.08	56	11.4%	0.95 [0.58, 1.32]
Droppelmann et al. 2020	-23.73	2.49	26	-23.24	2.92	41	8.9%	-0.18 [-0.67, 0.32]
Jotangia et al. 2015	-13.11	11.38	18	-14.1	8.07	20	6.7%	0.10 [-0.54, 0.74]
Redondo et al. 2012	-20	5.81	17	-32.64	6.38	11		Not estimable
Rees-Jones et al. 2012	-12.3	10.1	67	-14.72	10.43	54	11.6%	0.23 [-0.13, 0.59]
Yip et al. 2013	-14.6	11.22	30	-21.34	14.67	29	8.5%	0.51 [-0.01, 1.03]
Young et al. 2010	-11.36	10.54	22	-15.1	12.1	10	5.4%	0.33 [-0.42, 1.08]
Young et al. 2012	-8.5	9.56	16	-16.2	13.76	15	5.7%	0.64 [-0.09, 1.36]
Young et al. 2017	-9.24	1.52	33	-10.29	2.38	35	9.1%	0.52 [0.03, 1.00]
Total (95% CI)			**474**			**399**	**100.0%**	**0.30 [0.09, 0.51]**

Heterogeneity: Tau² = 0.06; Chi² = 21.87, df = 10 (P = 0.02); I² = 54%
Test for overall effect: Z = 2.82 (P = 0.005)

Fig. 5. Forest plot for violence and aggressiveness outcome. *Note*: Outlier excluded (i.e., Redondo et al., 2012) of this analysis.
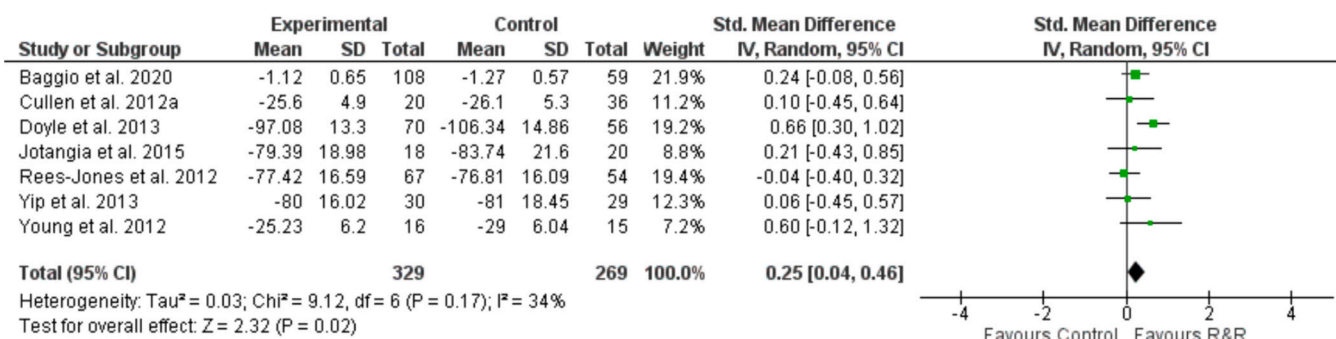
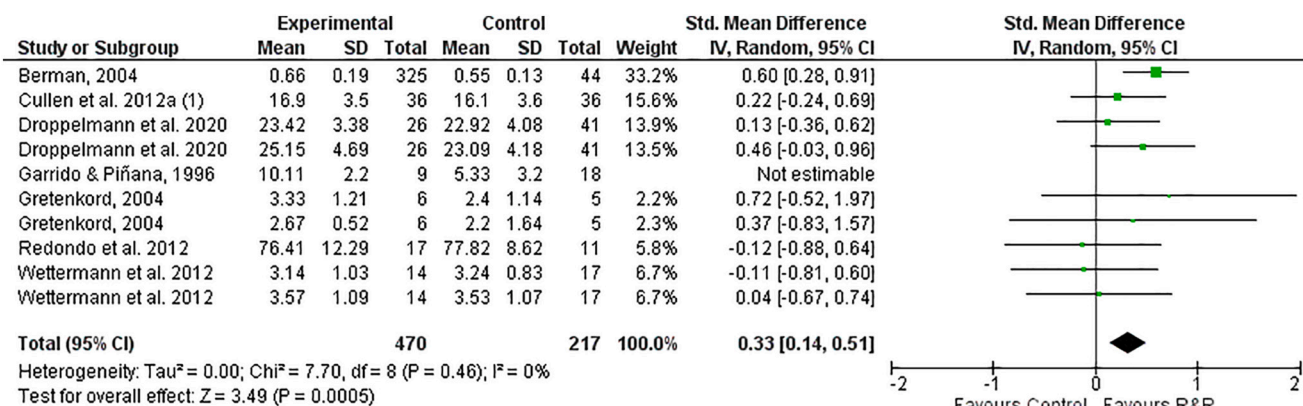**Fig. 6.** Forest plot for anger and hostility outcomes.

Fig. 7. Forest plot for empathy and social perspective taking outcome. *Note*: Outlier excluded (i.e., Garrido Genovés & Piñana, 1996) of this analysis.
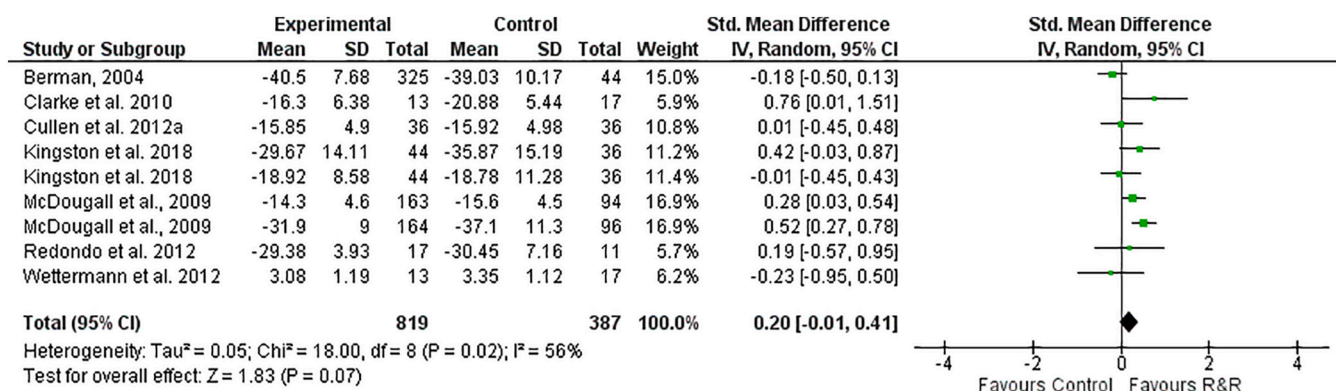

Fig. 8. Forest plot for criminal attitudes & sentiments outcome.

However, R&R, compared with other cognitive behavioral programs, showed no significant advantage (Lipsey et al., 2007).

Thus, the results of this review and previous reviews show that compared to wait-list or usual service, R&R appears to have a beneficial effect. Furthermore, the levels of heterogeneity in three outcomes (i.e., social problem solving, violence/aggressiveness, social perspective taking/empathy, and criminal attitudes) were significantly high but the effect was not moderated by the study design, the setting nor the authors' affiliation to the program. Future studies should examine the role of different factors which have been associated with the effectiveness of the programs such as demographic characteristics, facilitator skills and training, prison environment, levels of motivation, and levels of risk of reoffending (Auty & Liebling, 2019; Olver et al., 2011).

The program did not have an effect on criminal attitudes. This finding was unexpected, and a possible explanation is the measure used

in one study (Wettermann et al., 2012) which was based on interviews assessing prosocial skills, and unexpectedly the control group showed higher prosocial skills than the treated group. This might be explained by deceit and/or manipulation by the control group. When this study was excluded from the analysis, the effect was significantly positive for the treated group, that is, the program significantly reduced the criminal attitudes of the participants.

Although RCTs are considered the "gold standard" method to measure the impact of programs with individuals who commit offenses, there are potential issues owing to conflicts between the demands of research design and the realities of practice (Hollin & Palmer, 2009). There has been a debate about whether there is difference between RCTs and quasi-experimental designs (see Hollin & Palmer, 2009). However, it has been suggested that a high-quality quasi-experimental design is a satisfactory alternative when randomization is not possible (Farrington
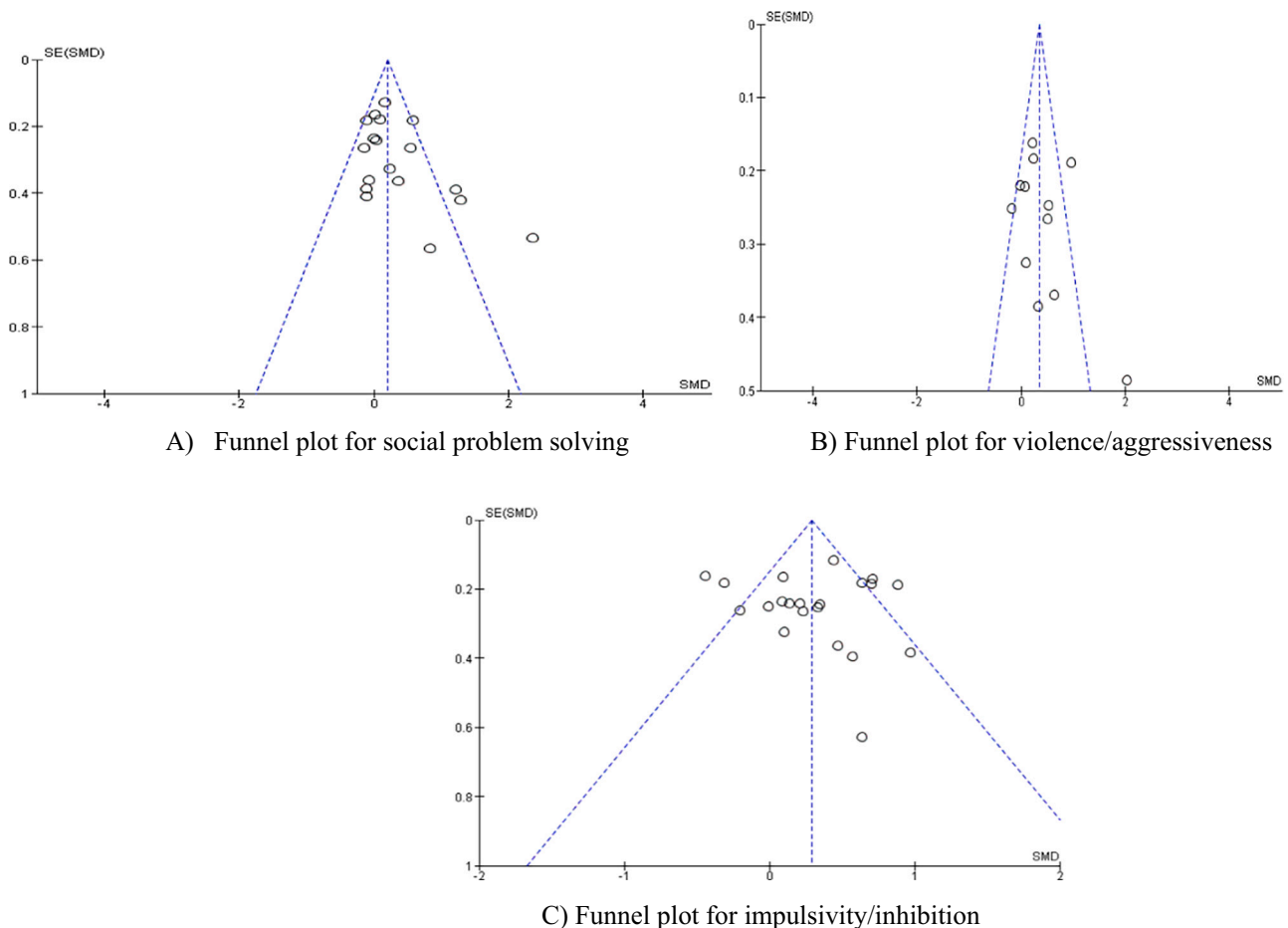
A) Funnel plot for social problem solving

B) Funnel plot for violence/aggressiveness

C) Funnel plot for impulsivity/inhibition

Fig. 9. Publication bias analysis for outcomes with k ≥ 10 based on the fixed-effect model. A) Funnel plot for social problem solving B) Funnel plot for violence/aggressiveness C) Funnel plot for impulsivity/inhibition.

& Jolliffe, 2002). In this study, the design did not have an impact on the results reported.

It is noteworthy that two studies from the USA and Chile reported that some null findings might be owing to the instrument used to measure outcomes (Droppelmann et al., 2020; Van Voorhis et al., 2001a, 2001b). Both studies reported that the psychometric properties of the scales (Chronbach's alpha) were weak, and differences across groups were minimal (Droppelmann et al., 2020; Van Voorhis et al., 2001a, 2001b). Authors speculated that either the instruments were not widely tested or validated, or the instrument simply may not be sensitive to change, and the tests may not be adequately corresponding to the aims of the intervention. However, both studies reported positive effects in the reduction of recidivism. Another study (Berman, 2004) stated that "*tests assessing the direct acquisition of cognitive and social skills, such as those recommended by program authors* (Ross & Fabiano, 1985)*, were not included*" (p. 89). Furthermore, Pullen (1996) reported different results depending on the measures used, and Cornet et al. (2016) included a limitation about the reliability and sensitivity of some measures, which might have reduced the significance of some scores. Therefore, future studies should not only explore which outcomes of the program are related to recidivism but also aim to develop an assessment tool that practitioners can use to measure the outcomes and reduce the variability of outcomes and scales used across studies, making the results of the intervention comparable.

4.1. Practical implications

Research shows that some factors including impulsivity, substance

abuse, criminal attitudes, antisocial personality disorder, antisocial peers, employment, and accommodation status correlate with crime (Bonta & Andrews, 2017; Ellis et al., 2019). Moreover, research highlights the importance of applying and tailoring interventions both in prison and in the community to reduce reoffending. However, the black box of the programs, that is, for whom and how they work is not well understood mainly because most of the meta-analytic reviews on program effectiveness for individuals who commit crime are focused on reoffending.

Although our results must be cautiously considered, the evidence reported in this review suggests that the R&R is effective in reducing violence, anger and impulsivity and improving (social) problem solving and perspective taking/empathy. It is important to note that these results are mainly applicable to the UK, where most of the participants were selected. However, this fact should encourage scholars to conduct evaluations in other countries where the program is implemented such as Canada, India, Mexico, and the US; and in the case of Spain, re-evaluate the program since it was evaluated in 1996.

4.2. Limitations

There are several limitations in this review. First, the number of studies included is relatively small, especially medium- and long-term studies and studies comparing R&R with an inactive control group. Despite the worldwide implementation of the R&R program (and its derivatives) to reduce recidivism, we were not able to find enough studies to conduct a meta-analytical analysis on some other outcomes related to offending and recidivism such as executive functioning

(Sánchez de Ribera et al., 2022; Spenser et al., 2019), and subgroup analyses based on participants' characteristics and program fidelity were not possible due to lack of variability across the studies and information, respectively. More data on specific outcomes would have enabled more robust findings and details on what works best in certain contexts. Additionally, quasi-experimental studies might be affected by confounding factors, increasing the likelihood of alternative explanations for observed effects (Sherman et al., 2002), and some biases were reported across the experimental studies, especially allocation sequence and concealment, incomplete data, contamination protection, and selective outcome reporting. Moreover, our study selection was also limited to studies written in English and Spanish language, and we did not include books and book chapters. Therefore, we do not rule out the possibility that studies published in this format or in other languages were missed, and might be the cause for the publication bias reported in this study. However, of the 20 included studies in the meta-analysis, only one was conducted in a non-Western population (Anonymous for review, 2020). Whether these findings are generalizable to Asian, African and Latin-American offenders remains to be documented.

5. Conclusion

Our findings suggest that the R&R program (and its derivatives) is effective in improving some psychosocial skills, and in turn, it can reduce recidivism. However, questions still remain (i.e., the medium-long term effects on some outcomes, the effect on different types of offenders, and different comparison groups) due to the small number of studies included. Additionally, we must be cautious owing to several limitations, which were mainly related to risk of bias and the diversity of outcomes reported across the studies. Future studies should examine the role of moderators (i.e., contextual factors) on the underlying mechanisms of the program. Additionally, evaluations need to draw on shared outcomes using common measures to corroborate the validity of these results.

CRediT authorship contribution statement

Olga Sánchez de Ribera: Conceptualization, Data curation, Formal analysis, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. **Violeta Chitgian Urzúa:** Data curation, Validation. **Genée Pienaar:** Formal analysis, Writing – review & editing.

Declaration of competing interest

The author(s) declared the following potential conflicts of interest with respect to the research, authorship, and/or publication of this article: The first author is a certified instructor of the R&R2 program. There are no other conflicts of interest.

Data availability

Data will be made available on request.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.avb.2024.101950.

References[*,†]

Antonowicz. (2005). The reasoning and rehabilitation program: Outcome evaluations with offenders. In M. McMurran, & J. McGuire (Eds.), *Social problem solving and offending: Evidence, evaluation and evolution* (pp. 163–167). Wiley.

Antonowicz, D. H. (2008). *The reasoning and rehabilitation program 20 years later.* Wilfrid Laurier University.

Aos, S., Phipps, P., Barnoski, R., & Lieb, R. (2001). *The comparative costs and benefits of programs to reduce crime.* Washington State Institute for Public Policy.

Auty, K., & Liebling, A. (2019). Exploring the relationship between prison social climate and reoffending. *Justice Quarterly, 37*(2), 1–24. https://doi.org/10.1080/07418825.2018.1538421

†Baggio, S., Weber, M., Rossegger, A., Endrass, J., Heller, P., Schneeberger, A., … Liebrenz, M. (2020). Reducing recidivism using the reasoning and rehabilitation program: A pilot multi-site-controlled trial among prisoners in Switzerland. *International Journal of Public Health, 65*(6), 801–810. https://doi.org/10.1007/s00038-020-01372-9

Bandura, A. J. (1977). *Social learning theory.* Prentice Hall.

†Berman, A. H. (2004). The Reasoning and Rehabilitation Program: Assessing short- and long-term outcomes among male Swedish prisoners. *Journal of Offender Rehabilitation, 40*(1–2), 85–103. https://doi.org/10.1300/J076v40n01_05

*Bonta, J., & Andrews, D. A. (2017). *The psychology of criminal conduct* (7th ed.). Routledge.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (Eds.). (2009). *Introduction to meta-analysis.* Wiley. https://doi.org/10.1002/9780470743386.

†Clarke, A. Y., Cullen, A. E., Walwyn, R., & Fahy, T. (2010). A quasi-experimental pilot study of the reasoning and rehabilitation programme with mentally disordered offenders. *Journal of Forensic Psychiatry & Psychology, 21*(4), 490–500. https://doi.org/10.1080/1478994 0903236391

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Academic Press.

†Cornet, Laan, P. H.v.d., Nijman, H. L., Tollenaar, N., & Kogel, C. H.d. (2016). Does a cognitive skills training program for prisoners affect neurocognitive functioning and heart rate activity? *Criminal Justice and Behavior, 43*(11), 1481–1504. https://doi.org/10.1177/00938548 16643732

†Cullen, A. E., Clarke, A. Y., Kuipers, E., Hodgins, S., Dean, K., & Fahy, T. (2012a). A multi-site randomized controlled trial of a cognitive skills programme for male mentally disordered offenders: Social-cognitive outcomes. *Psychological Medicine, 42*(3), 557–569. https://doi.org/10.1017/S0033291711001553

†Cullen, A. E., Clarke, A. Y., Kuipers, E., Hodgins, S., Dean, K., & Fahy, T. (2012b). A multisite randomized trial of a cognitive skills program for male mentally disordered offenders: Violence and antisocial behavior outcomes. *Journal of Consulting and Clinical Psychology, 80*(6), 1114–1120. https://doi.org/10.1037/a0030291

†Curran, & Bull, R. (2009). Ross programme: Effectiveness with young people in residential childcare. *Psychiatry, Psychology and Law, 16*(1), S81–S89. https://doi.org/10.1080/1321871080 2242029

Deeks, J.J., Higgins, J.P., & Altman, D.G. (2023). Chapter 9: Analyzing data and undertaking meta-analyses. In J.P.T Higgins & Thomas, J. (Senior Eds.). Cochrane handbook for systematic reviews of interventions version 6.4. Cochrane. Available from www.training.cochrane.org/handbook.

Deeks J.J., Higgins J.P.T., Altman D.G. (editors). Chapter 10: Analysing data and undertaking meta-analyses. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, Welch VA (editors). Cochrane Handbook for Systematic Reviews of Interventions version 6.4 (updated August 2023). Cochrane, 2023. Available from www.training.cochrane.org/handbook.

†Doyle, M., Khanna, T., Lennox, C., Shaw, J., Hayes, A., Taylor, J., Roberts, A., & Dolan, M. (2013). The effectiveness of an enhanced thinking skills programme in offenders with antisocial personality traits. *The Journal of Forensic Psychiatry & Psychology, 24*(1), 1–15. https://doi.org/10.1080/14789949.2012.752519

†Droppelmann, C., Osorio, V., & Jara, D. (2020). *Adaptación, implementación y evaluación de impacto del programa de intervención cognitivo conductual "Reasoning & Rehabilitation" para la población infractora adulta de centros privativos de libertad [Adaptation, implementation and impact evaluation of the cognitive behavioral intervention program "Reasoning & Rehabilitation" for the adult offender population in prisons] Centros de Estudios Justicia & Sociedad.* Pontificia Universidad Católica de Chile.

Effective Practice and Organisation of Care (EPOC). (2014). Suggested risk of bias criteria for EPOC reviews Oslo: Norwegian Knowledge Centre for the Health Services. Available: https://epoc.cochrane.org/epoc-specific-resources-review-authors.

Ellis, L., Farrington, D. P., & Hoskin, A. W. (2019). *Handbook of crime correlates.* Academic Press.

Farrington, D. P., & Jolliffe, D. (2002). *A feasibility study into using a randomised controlled trial to evaluate treatment pilots at HMP Whitemoor.* Home Office Online Report 14/02. London: Home Office

---

[*] Studies included in the systematic review.

[†] Studies included in the systematic review and meta-analysis.

Gaffney, H., Ttofi, M. M., & Farrington, D. P. (2021). Effectiveness of school-based programs to reduce bullying perpetration and victimization: An updated systematic review and meta-analysis. *Campbell Systematic Reviews, 17*(2), Article e1143. https://doi.org/10.1002/cl2.1143

Galway, R., Swales, M. A., & Wane, J. (2022). The efficacy of offence-specific interventions in reducing risk and recidivism in women: A systematic review. *The Journal of Forensic Psychiatry & Psychology, 33*(3), 291–322. https://doi.org/10.1080/14789949.2022.2053185

†Garrido Genovés, V., & Piñana, A. M. (1996). El modelo cognitivo aplicado a delincuentes institucionales: El pensamiento prosocial. *Revista Complutense de Educación, 7*(2), 137. https://revistas.ucm.es/index.php/RCED/article/view/RCED9696220137A.

*Garrido, V., & Sanchis, J. R. (1991). The cognitive model in the treatment of Spanish offenders: Theory and practice. *Journal of Correctional Education, 42*(2), 111–118. https://www.jstor.org/stable/41970876.

†Gretenkord, L. (2004). *R&R treatment effects: Haina pilot study* (Germany: Unpublished report prepared for the Haina Forensic Psychiatric Hospital).

Heeks, M., Reed, S., Tafsiri, M., & Prince, S. (2018). *The economic and social costs of crime* (2nd ed.). London: Home Office. Research Report 99. Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/954485/the-economic-and-social-costs-of-crime-horr99.pdf.

Higgins, J. P., Altman, D. G., Gøtzsche, P. C., Jüni, P., Moher, D., Oxman, A. D., … Cochrane Statistical Methods Group. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *British Medical Journal (Clinical Research Ed.), 343*, Article d5928. https://doi.org/10.1136/bmj.d5928

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*, 557–560. Cochrane. Available from www.training.cochrane.org/handbook.

Hollin, C. R., & Palmer, E. J. (2009). Cognitive skills programmes for offenders. *Psychology, Crime & Law, 15*(2), 147–164. https://doi.org/10.1080/10683160802190871

Jackson, D., & Turner, R. (2017). Power analysis for random-effects meta-analysis. *Research Synthesis Methods, 8*(3), 290–302. https://doi.org/10.1002/jrsm.1240

Jaitman, L. (2019). Frontiers in the economics of crime: Lessons for Latin America and the Caribbean. *Latin America Economic Review, 28*(19), 1–36. https://doi.org/10.1186/s40503-019-0081-5

James, A. C., Reardon, T., Soler, A., James, G., & Creswell, C. (2020). Cognitive behavioural therapy for anxiety disorders in children and adolescents. *Cochrane Database of Systematic Reviews, 11*, Article CD013162. https://doi.org/10.1002/14651858.CD013162.pub2

†Jotangia, A., Rees-Jones, A., Gudjonsson, G. H., & Young, S. (2015). A multi-site controlled trial of the R&R2MHP cognitive skills program for mentally disordered female offenders. *International Journal of Offender Therapyx and Comparative Criminology, 59*(5), 539–559. https://doi.org/10.1177/0306624X13512092

*Khodayarifard, M., Shokoohi-Yekta, M., & Hamot, G. E. (2010). Effects of individual and group cognitive-behavioral therapy for male prisoners in Iran. *International Journal of Offender Therapy and Comparative Criminology, 54*(5), 743–755. https://doi.org/10.1177/030662 4X09344840

Kim, B., Merlo, A. V., & Benekos, P. J. (2013). Effective correctional intervention programmes for juveniles: Review and synthesis of meta-analytic evidence. *International Journal of Police Science & Management, 15*(3), 169–189. https://doi.org/10.1350/ijps.2013.15.3.310

†Kingston, D. A., Olver, M. E., McDonald, J., & Cameron, C. (2018). A randomised controlled trial of a cognitive skills programme for offenders with mental illness. *Criminal behaviour and mental health: CBMH, 28*(4), 369–382. https://doi.org/10.1002/cbm.2077

Lipsey, M. W., Landenberger, N. A., & Wilson, S. J. (2007). Effects of cognitive-behavioral programs for criminal offenders. *Campbell Systematic Reviews, 3*(1), 1–27. https://doi.org/10.4073/csr.2007.6

†McDougall, C., Perry, A. E., Clarbour, J., Bowles, R., & Worthy, G. (2009). *Evaluation of HM prison service enhanced thinking skills programme: Report on the outcomes from a randomised controlled trial*. Ministry of Justice Research Series 3/09. London: Ministry of Justice.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & the PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151*, 264–269. https://doi.org/10.7326/0003-4819-151-4-200908180-00135v

Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2011). A meta-analysis of predictors of offender treatment attrition and its relationship to recidivism. *Journal of Consulting and Clinical Psychology, 79*(1), 6–21. https://doi.org/10.1037/a0022200

Page, M. J., Higgins, J. P. T., & Sterne, J. A. C. (2023). Chapter 13: Assessing risk of bias due to missing results in a synthesis. Senior Eds. In J. P. T. Higgins, & J. Thomas (Eds.), *Cochrane handbook for systematic reviews of interventions version 6.4*. Cochrane. Available from www.training.cochrane.org/handbook.

Pearson, F. S., Lipton, D. S., Cleland, C. M., & Yee, D. S. (2002). The effects of behavioral/cognitive-behavioral programs on recidivism. *Crime & Delinquency, 48*(3), 476–496. https://doi.org/10.1177/0011128702048003006

*Pullen, S. (1996). *Evaluation of the reasoning and rehabilitation cognitive skills development program as implemented in Juvenile ISP in Colorado*. Unpublished report. Colorado Division of Criminal Justice.

†Redondo, S., Martínez-Catena, A., & Andrés-Pueyo, A. (2012). Therapeutic effects of a cognitive-behavioral treatment with juvenile offenders. *The European Journal of Psychology Applied to Legal Context, 4*(2), 159–178.

Redondo, S., Sanchez-Meca, J., & Garrido, V. (1999). Crime treatment in Europe: A review of outcome studies. In J. McGuire (Ed.), *Offender rehabilitation and treatment:*

Effective programs and policies to reduce re-offending (pp. 113–141). John Wiley and Sons.

†Rees-Jones, A., Gudjonsson, G., & Young, S. (2012). A multi-site controlled trial of a cognitive skills program for mentally disordered offenders. *BMC Psychiatry, 12*(44), 1–11. https://doi.org/10.1186/1471-244X-12-44

ReviewManager 5 (RevMan 5). (2014). Copenhagen: NordicCochrane Centre, The Cochrane Collaboration. Version 5.3.

Robinson, D., & Porporino, F. (2001). Programming in cognitive skills: The reasoning and rehabilitation programme. In C. R. Hollin (Ed.), *Handbook of offender assessment and treatment* (pp. 179–193). Wiley.

Ross, R. R., & Fabiano, E. A. (1985). *Time to think: A cognitive model of delinquency prevention and offender rehabilitation*. Johnson City, TN: Institute of Social Sciences and Arts.

Ross, R. R., Fabiano, E. A., & Ewles, C. D. (1988). Reasoning and rehabilitation. *International Journal of Offender Therapy and Comparative Criminology, 32*, 29–35.

Ross, R. R., & Hilborn, J. (2008). *Rehabilitating rehabilitation: Neurocriminology for prevention and treatment of antisocial behavior*. Ottawa: Cognitive Centre of Canada. www.cognitivecentre.ca.

†Sánchez de Ribera, O. (2015). *Neuropsychological functions in sex offenders: Empirical and relations and an evaluation of the thinking skills programme* (Doctoral thesis). The University of Cambridge https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.708857.

Sánchez de Ribera, O., Trajtenberg, N., & Cook, S. (2022). Executive functioning among first time and recidivist inmates in Uruguay. *Applied Neuropsychology: Adult, 29*(5), 1242–1249. https://doi.org/10.1080/23279095.2020.1864634

Schmucker, M., & Lösel, F. (2017). Sexual offender treatment for reducing recidivism among convicted sex offenders: A systematic review and meta-analysis. *Campbell Systematic Reviews, 13*(1), 1–75. https://doi.org/10.4073/csr.2017.8

Sherman, L. W., Farrington, D., Welsh, B. C., & Mackenzie, D. L. (2002). *Evidence-based crime prevention* (Revised ed.). Routledge.

Spenser, K. A., Bull, R., Betts, L., & Winder, B. (2019). Executive functioning as a predictive measure of offending behaviour. *Journal of Criminal Psychology, 9*(1), 10–22. https://doi.org/10.1108/JCP-07-2018-0032

Sterne, J. A., Egger, M., Moher, D., & Boutron, I. (2017). Chapter 10: Addressing reporting biases. In J. P. T. Higgins, R. Churchill, J. Chandler, & M. S. Cumpston (Eds.), *Cochrane handbook for systematic reviews of interventions version 5.2.0*. Cochrane. Available from www.training.cochrane.org/handbook.

Tanriver-Ayder, E., Faes, C., van de Casteele, T., McCann, S. K., & Macleod, M. R. (2021). Comparison of commonly used methods in random effects meta-analysis: Application to preclinical data in drug discovery research. *British Medical Journal Open Science, 5*(1), Article e100074. https://doi.org/10.1136/bmjos-2020-100074

The Jamovi Project. (2021). Jamovi. (Version 2.2) [Computer software]. Retrieved from https://www.jamovi.org/.

Tong, L. S., & Farrington, D. P. (2006). How effective is the "reasoning and rehabilitation" programme in reducing reoffending? A meta-analysis of evaluations in four countries. *Psychology, Crime and Law, 12*(1), 3–24. https://doi.org/10.1080/10683160512331316253

*Van Voorhis, P., Spruance, L., Johnson-Listwan, S., Ritchey, P. N., Pealer, J., & Seabrook, R. (2001a). The Georgia cognitive skills experiment, outcome evaluation: Phase one. Retrieved from https://www.uc.edu/content/dam/uc/ccjr/docs/reports/project_reports/Georgia_Phase_1_final_report.pdf.

*Van Voorhis, P., Spruance, L. M., Ritchey, P. N., Johnson-Listwan, S., Seabrook, R., & Pealer, J. (2001b). The Georgia cognitive skills experiment, outcome evaluation: Phase II. Retrieved from https://www.uc.edu/content/dam/uc/ccjr/docs/reports/project_reports/Georgia_Phase_II_final_report.pdf.

Walters, G. D. (2022). Crime and social cognition: A meta-analytic review of the developmental roots of adult criminal thinking. *Journal of Experimental Criminology, 18*, 183–207. https://doi.org/10.1007/s11292-020-09435-w

†Wettermann, A., Schläfke, D., & Fegert, J. (2012). The modification of criminogenic factors on addicted offenders. The effectiveness of the reasoning and rehabilitation program. *International Journal of Law and Psychiatry, 35*(3), 202–206. https://doi.org/10.1016/j.ijlp. 2012.02.009

†Wettermann, A., Völlm, B., & Schläfke, D. (2020). Highly structured treatment programs for addicted offenders: Comparing the effects of the reasoning & rehabilitation program and DBT-F. *Frontiers in Psychiatry, 11*, Article 499241. https://doi.org/10.3389/fpsyt.2020.499241

Wilson, D. B., Bouffard, L. A., & Mackenzie, D. L. (2005). A quantitative review of structured, group-oriented, cognitive-behavioral programs for offenders. *Criminal Justice and Behavior, 32*(2), 172–204. https://doi.org/10.1177/0093854804272889

†Yip, V., Gudjonsson, G. H., Perkins, D., Doidge, A., Hopkin, G., & Young, S. (2013). A non-randomised controlled trial of the R&R2MHP cognitive skills program in high risk male offenders with severe mental illness. *BMC Psychiatry, 13*(267), 1–11. https://doi.org/10.1186/1471-244X-13-267

†Young, S., Chick, K., & Gudjonsson, G. (2010). A preliminary evaluation of reasoning and rehabilitation 2 in mentally disordered offenders (R&R2M) across two secure forensic settings in the United Kingdom. *Journal of Forensic Psychiatry & Psychology, 1-14*. https://doi.org/10.1080/14789940903513203

†Young, S., Emilsson, B., Sigurdsson, J. F., Khondoker, M., Philipp-Wiegmann, F., Baldursson, G., … Gudjonsson, G. (2017). A randomized controlled trial reporting functional outcomes of cognitive-behavioural therapy in medication-treated adults with ADHD and comorbid psychopathology. *European Archives of Psychiatry and Clinical Neuroscience, 267*(3), 267–276. https://doi.org/10.1007/s00406-016-0735-0

†Young, S., Hopkin, G., Perkins, D., Farr, C., Doidge, A., & Gudjonsson, G. (2012). A controlled trial of a cognitive skills program for personality-disordered offenders. *Journal of Attention Disorders*, 1–10. https://doi.org/10.1177/1087054711 430333

†Young, S., Khondoker, M., Emilsson, B., Sigurdsson, J. F., Philipp-Wiegmann, F., Baldursson, G., … Gudjonsson, G. (2015). Cognitive–behavioural therapy in medication-treated adults with attention-deficit/hyperactivity disorder and co-morbid psychopathology: A randomized controlled trial using multi-level analysis. *Psychological Medicine, 45*(13), 2793–2804. https://doi.org/10.1017/S0033291715000756

Yukhnenko, D., Sridhar, S., & Fazel, S. (2019). A systematic review of criminal recidivism rates worldwide: 3-year update. *Wellcome Open Research, 4*, 28. https://doi.org/10.12688/wellcomeopenres.14970.3