CBT Skills Programmes (Prevention / Universal & At-risk)

Toolkit technical report

Ciara Keenan, Iman Idrees, Lauren Murphy, Karen Allen, Jessica Hinks, Ofelia Torres, Jenny Towers, Jaimee S. Mallion, Sophie Hedges, Ella James-Brabham

June 2026

About National Children's Bureau

This report has been produced by the National Children's Bureau on behalf of the Youth Endowment Fund. The National Children's Bureau works collaboratively across the issues affecting children to influence policy and get services working together to deliver a better childhood. They were commissioned by the Youth Endowment Fund (YEF) as their Toolkit Partner 2023-2026. The Toolkit partnership is managed on behalf of NCB by Dr Ciara Keenan. Any queries relating to the methods should be directed to ckeenan@ncb.org.uk

About Youth Endowment Fund

The Youth Endowment Fund's mission is to prevent children and young people becoming involved in violence. They do this by finding out what works and building a movement to put this knowledge into practice. The fund was established in March 2019 by children's charity Impetus, with a £200m endowment and ten-year mandate from the Home Office. For more information, please visit www.youthendowmentfund.org.uk.

Acknowledgements

Our deepest appreciation goes to all those who shared their insights and expertise to enhance the quality of this research. Special acknowledgement is due to Dr Laura Knight, Head of Toolkit, Nigel Tungwarara and Dr Chad Hemady, Senior Research Managers at the Youth Endowment Fund, for their diligent work and unwavering support throughout this project. Our sincere appreciation also goes to Professor James Thomas and the entire team at the EPPI-Centre, whose dedication and countless hours have been instrumental in the development of this work.

Abstract/Plain Language Summary

The primary objective of this report is to review the evidence on the effectiveness of **CBT Skills Programmes** as a strategy to prevent violence, crime and offending among children and young people. CBT Skills Programmes are preventative interventions that provide psychoeducation and use cognitive behavioural techniques to teach practical skills such as emotional regulation, anger management, self-control, perspective-taking, and effective and assertive communication.

These programmes are typically collaborative and goal-focused. They are usually manualised interventions delivered over a fixed-duration in non-clinical settings such as schools, youth clubs, community centres, youth justice settings, and home-based contexts. They are usually facilitated by trained practitioners, such as teachers, youth workers, or other non-clinical staff, rather than licensed mental health professionals. However, they can be delivered by licensed professionals in a non-clinical settingCBT Skills Programmes are generally aimed at the general population, or at children and young people who are at risk of involvement in violence or offending, or who are already displaying behavioural difficulties, rather than those receiving formal clinical treatment for a mental health condition.

Key findings

- CBT Skills programmes appears to reduce violence by 72%, based on a meta-analysis of 56 measured outcomes across 22 studies. There is an overall evidence security rating of Level 3 (Moderate) for this finding.
- CBT Skills programmes might reduce crime and offending by 71%, based on a meta-analysis of 11 measured outcomes across four studies. There is an overall evidence security rating of Level 1 (Very Low) for this finding.
- Across all related outcomes, CBT Skills programmes are estimated to have an overall beneficial effect (g = -0.58) equating to a High YEF impact rating, based on a meta-analysis of 251 measured outcomes across 37 studies.

There is an overall evidence security rating of Level 3 (Moderate) for this finding.

- Subgroup analyses suggested that CBT Skills programmes were associated with beneficial effects across a wide range of populations, settings, and delivery formats. The clearest and most robust evidence of benefit was found in studies involving children and young people with externalising or behavioural dysregulation difficulties and in interventions delivered in school or education settings.
- There was little consistent evidence that programme intensity, delivery format, age group, or gender composition reliably explained differences in effects. Although some subgroup differences were statistically significant, many estimates were based on small numbers of studies and should therefore be interpreted cautiously.
- Acceptability of CBT Skills programmes was usually positive when interventions felt engaging, practical and supportive, but the evidence base was limited and sometimes relied more on teacher or parent views than on children and young people's own perspectives.
- CBT Skills programmes appeared easier to implement when they were adapted to the developmental needs of children and young people, for example through shorter sessions, interactive activities, and more accessible materials.
- Fidelity was generally higher in studies that used manuals, facilitator training, regular supervision, and direct monitoring methods such as recordings, observations, or adherence checklists.
- Implementation was stronger when programmes were designed to fit the practical realities of delivery settings, particularly schools, by aligning with timetables, staffing capacity, and everyday routines.

Conclusion

Our meta-analysis finds that CBT Skills programmes reduce violence and crime-related outcomes among children and young people and are estimated to have beneficial effects across a broader range of behavioural, emotional, and social

outcomes overall. Acceptability was generally positive when programmes were experienced as engaging, practical, and supportive, although the evidence base was limited and often drew on a mix of perspectives from teachers, parents, and children and young people. Implementation appeared strongest when programmes were adapted to developmental needs, supported by manuals, training, supervision and fidelity monitoring, and designed to fit the practical realities of delivery settings, particularly schools.

Table of Contents

List of Tables

List of Figures

Preface on Terminology

This review draws on evidence spanning over half a century, during which language around personal characteristics has evolved significantly. At times, we may have to reproduce original terminology used in studies which we recognise today as being outdated and unacceptable offensive terms. This only occurs when the terminology is used in direct quotations or refers to an outcome that the authors measured that remains relevant to our analysis. The wider narrative will adhere to current inclusive-language standards guided by the National Children's Bureau, Youth Endowment Fund, and Race Equality Foundation. These guiding principles include using capitalisation to acknowledge shared identities (e.g., Black, Asian), whilst not capitalising white due to its association with white supremacy. The review also avoids deficit framing and respects individuals' self-identification. Person-first language will generally be used when referring to children and young people, except for Deaf and autistic communities, who widely prefer identity-first language. The team acknowledges limitations in terminology and strives for respectful and precise representation throughout. The full preface on terminology can be accessed here.

Objective and Approach

The objective of this report is to review the evidence on the effectiveness of **CBT Skills Programmes** in reducing and/or preventing violence and offending involving children and young people. CBT Skills-based programmes use cognitive behavioural techniques to support children and young people to develop practical skills linked to improved self-regulation and more prosocial behaviour. These collaborative, goal-focussed programmes typically focus on areas such as emotional regulation, anger management, social problem-solving, perspective-taking, and impulse control. Rather than providing high-intensity clinical treatment (which is the focus of the Clinical CBT strand), they are designed to provide psycho-education and collaboratively teach concrete cognitive and behavioural skills that children and young people can apply in everyday situations.

Included programmes are delivered to children and young people aged 0 to 17 years, either universally or to groups considered at-risk, including those with behavioural difficulties or those at risk of, or already involved in, offending. In contrast to Clinical CBT, these interventions do not require a formal mental health condition or neurodevelopmental disorder diagnosis, and are not primarily intended to treat psychiatric or trauma-related conditions. They are typically delivered in non-clinical settings, such as schools, community centres, youth clubs, youth justice settings, or home-based environments. Included programmes are often delivered by trained practitioners including teachers, youth workers, and other non-clinical staff, as well as clinical staff delivering the intervention in a non-clinical, programme-based format.

This technical report draws on a comprehensive systematic review methodology and includes:

- **37 effectiveness studies** provided outcome data for meta-analysis, including two from the UK and 35 from 10 other countries. These studies contributed 56 violence outcomes across 22 studies, 11 crime and offending outcomes across four studies, and 251 outcomes across the full YEF Outcomes Framework.
- **Thirteen studies providing implementation insights.** Findings are reported according to Proctor's (2011) framework.

The remainder of this report is structured as follows: First, the **Description of the Intervention** outlines the key components of CBT Skills and its intended implementation. Second, **How Effective is the Intervention?** presents findings from our meta-analysis on violence and crime reduction and broader social outcomes. Third, **Who Does it Work For?** examines evidence on the populations that benefit most from CBT Skills. Fourth, **What Factors Affect Implementation?** explores key factors which influence the implementation of CBT Skills interventions. Fifth, **How Much Does It Cost?** reviews available cost data. Finally, the **Conclusion and Takeaway Messages** summarise key findings and recommendations, followed by **Appendices** detailing the systematic review methodology and characteristics of included research.

Description of the Intervention

In the following section details are provided on the interventions which inform this report, noting their key components, any equipment, materials, supplies or training required, the duration and intensity of interventions, who delivered the interventions, and where and how the interventions were delivered.

Features of the approach

Cognitive Behavioural Therapy (CBT) is a type of talking therapy that has been used to address a range of psychological difficulties. When applied to violence prevention, CBT is based on the idea that negative or impulsive thoughts and behaviours can influence a young person's emotional responses and increase the likelihood of lashing out or acting aggressively. A core principle of CBT is the interaction between thoughts, feelings, and behaviours, with beliefs shaping how situations are interpreted, emotions driving responses, and behaviours reinforcing patterns of thinking.

CBT aims to help children and young people become more aware of these negative thoughts and learn to change or manage them, often through a collaborative process. The therapist or facilitator might work with a child to explore how their assumptions relate to reality, develop a better understanding of other people's behaviour and motivations, and use problem-solving skills to cope with difficult situations and work towards the young person's goals.
CBT Skills programmes focus on cognitive skills training. They teach specific interpersonal problem-solving and self-regulation skills to a broad population, i.e., general population or at-risk young people. They typically utilise a fixed curriculum and are delivered in non-clinical environments (schools, community settings) facilitated by trained lay practitioners, such as teachers or youth workers, rather than licensed mental health clinicians. The interventions often take the form of group-based courses or curricula that teach cognitive and behavioural skills. Role-play, group discussion, and practice/ home tasks allow children to practice new skills (e.g. calmly resolving conflicts, considering others' perspectives) in a supportive peer environment.

Key components of CBT Skills

Most CBT Skills interventions were designed to support children and young people with behavioural difficulties or externalising behaviours, particularly anger and aggression (Augustyniak et al., 2009; Avci & Kelleci, 2016; Bowman & Auerbach,

1982; Cole et al., 2013; Daunic et al., 2006; Doğan & Çam, 2020; Etscheidt, 1991; Freiden, 2006; Glick & Goldstein, 1987; Karatas & Gokcakan, 2009; Kettlewell & Kausch, 1983; Kumuyi et al., 2022; Lesure-Lester, 2002; Lochman et al., 2017; Logsdon, 2003; Moynahan & Strømgren, 2005; Njardvik et al., 2022; Oparaduru, 2017; Poulin et al., 2001; Saba et al., 2023; Squires & Caddick, 2012; Sukhodolsky et al., 2000; Te Brinke et al., 2021; Watson, 2009; Yeo & Choi, 2011).

A smaller group of interventions targeted other behavioural or emotional concerns, including:

- Bullying reduction (Ime, 2024; McLaughlin, 2009)
- School attendance difficulties (Johnsen et al., 2024)
- Emotional wellbeing, resilience or anxiety prevention (Kozina, 2021; Orgilés et al., 2023; Schwartz-Mette et al., 2024)
- Substance use, self-harm and HIV-related risk (Esposito-Smythers et al., 2017; Wagner et al., 2014)
- Rehabilitation for young people who have offended (Pullen, 1996).

Programmes addressed these difficulties through core skills which were taught across interventions, in particular:

- **Problem-solving:** Many interventions taught children to identify problems, think before acting, consider consequences and evaluate alternative solutions (Augustyniak et al., 2009; Bowman & Auerbach, 1982; Cook et al., 2014; Daunic et al., 2006; Esposito-Smythers et al., 2017; Etscheidt, 1991; Glick & Goldstein, 1987; Hawkins et al., 1991; Kozina, 2021; Kumuyi et al., 2022; Lochman et al., 2017; Logsdon, 2003; Njardvik et al., 2022; Poulin et al., 2001; Pullen, 1996; Schwartz-Mette et al., 2024; Squires & Caddick, 2012; Te Brinke et al., 2021; Yeo & Choi, 2011).
- **Anger awareness and reduction:** This was taught through strategies such as identifying triggers, recognising signs of anger escalation, considering the consequences of anger and practising anger-management techniques (Augustyniak et al., 2009; Avci & Kelleci, 2016; Cole et al., 2013; Cook et al., 2014; Daunic et al., 2006; Doğan & Çam, 2020; Etscheidt, 1991; Freiden, 2006; Glick & Goldstein, 1987; Hawkins et al., 1991; Lesure-Lester, 2002; Logsdon, 2003; McLaughlin, 2009; Njardvik et al., 2022; Nwolisa, 2023; Saba et al., 2023; Squires & Caddick, 2012; Sukhodolsky et al., 2000; Te Brinke et al., 2021; Watson, 2009).

- **Understanding the relationship between thoughts, recognising, identifying and managing emotions and behaviours:** Many studies emphasised understanding how thoughts influence behaviour, including identifying negative attributions, cognitive distortions and catastrophising (Avci & Kelleci, 2016; Cook et al., 2014; Daunic et al., 2006; Freiden, 2006; Ime, 2024; Johnsen et al., 2024; Karatas & Gokcakan, 2009; Kumuyi et al., 2022; Logsdon, 2003; Njardvik et al., 2022; Nwolisa, 2023; Oparaduru, 2017; Pullen, 1996; Saba et al., 2023; Schwartz-Mette et al., 2024; Squires & Caddick, 2012; Sukhodolsky et al., 2000; Te Brinke et al., 2021; Yeo & Choi, 2011).
- **Coping strategies:** Programmes frequently taught relaxation and regulation strategies such as positive self-talk, breathing exercises, guided imagery and techniques to reduce rumination (Augustyniak et al., 2009; Avci & Kelleci, 2016; Bowman & Auerbach, 1982; Cole et al., 2013; Cook et al., 2014; Daunic et al., 2006; Doğan & Çam, 2020; Etscheidt, 1991; Glick & Goldstein, 1987; Ime, 2024; Kettlewell & Kausch, 1983; Kozina, 2021; Lesure-Lester, 2002; Njardvik et al., 2022; Nwolisa, 2023; Saba et al., 2023; Schwartz-Mette et al., 2024; Squires & Caddick, 2012; Sukhodolsky et al., 2000; Te Brinke et al., 2021; Watson, 2009).
- **Social skills, communication and empathy:** Many interventions included components focused on improving social and communication skills and developing empathy (Cook et al., 2014; Doğan & Çam, 2020; Esposito-Smythers et al., 2017; Freiden, 2006; Glick & Goldstein, 1987; Ime, 2024; Kozina, 2021; Logsdon, 2003; Poulin et al., 2001; Pullen, 1996; Schwartz-Mette et al., 2024).
- **Perspective-taking:** Several programmes encouraged children to understand other people's emotions and viewpoints (Augustyniak et al., 2009; Cook et al., 2014; Freiden, 2006; Ime, 2024; Kozina, 2021; Logsdon, 2003; McLaughlin, 2009; Te Brinke et al., 2021; Yeo & Choi, 2011).

Some programmes explicitly incorporated additional components such as goal setting (Burraston et al., 2012; Freiden, 2006; Hawkins et al., 1991; Lochman et al., 2017; Njardvik et al., 2022; Schwartz-Mette et al., 2024; Wagner et al., 2014), assertiveness (Avci & Kelleci, 2016; Esposito-Smythers et al., 2017; McLaughlin, 2009) and elements focused on moral reasoning, character development or

values education (Cook et al., 2014; Freiden, 2006; Glick & Goldstein, 1987; Pullen, 1996; Saba et al., 2023).

Skills were taught using a range of activities, including games, stories and video clips. These were reinforced through role-play, positive modelling, behaviour rehearsal and group discussions.

For example, in Avci & Kelleci (2016), early sessions used games, stories and scenarios to facilitate understanding and group bonding. Homework helped children reflect on real-life experiences of anger and its consequences. Children generated alternative responses to these situations, practised them through role-play and received feedback in later sessions. In Bowman & Auerbach (1982), the General Problem Solving Model (GPSM) guided children to identify problems, pause, think, plan and evaluate possible actions. Relaxation was promoted through positive statements, problem-solving exercises and breathing techniques.

In Burraston et al. (2012), young people examined their behaviours, attitudes and underlying beliefs, identifying which beliefs might be contributing to their difficulties. This work was supported through group discussions, exercises and video examples illustrating natural consequences of behaviour. The brief in person programme was supplemented with a 'Cell Phone' component, which provided daily check ins on progress toward goals and delivered supportive messages from significant others.

Several interventions utilised positive reinforcement and incentives to encourage desired behaviours. In the interventions described by Avci and Kelleci (2016), Ime (2024), Lochman et al. (2017) and Yeo and Choi (2011), children earned tokens or points that could be exchanged for prizes or group activities. Other programmes (Kettlewell & Kausch, 1983; Logsdon, 2003; Njardvik et al., 2022; Schwartz-Mette et al., 2024; Sukhodolsky et al., 2000) rewarded participants with items like ice cream, snacks or small toys for following rules, completing homework or showing positive behaviour. Kozina (2021) reported celebrating progress with an end-of-programme party. In contrast, the intervention in Oparaduru (2017) emphasised encouraging young people to focus on long-term rather than short-term rewards and values.

In addition to supporting children and young people, several interventions involved parents and carers. Two programmes (Lochman et al., 2017; Poulin et al., 2001) included dedicated parenting skills sessions. Esposito-Smythers et al. (2017) reported parental involvement through both joint and separate parent/teen interaction exercises. Johnsen et al. (2024) similarly reported sessions with the young person and parents together, alongside two additional parent-only sessions.

In other interventions, parents and carers were kept informed about what their children were learning. This was done through parent information sessions (Cole et al., 2013; Kozina, 2021) or through weekly letters sent home (Njardvik et al., 2022; Orgilés et al., 2023; Schwartz-Mette et al., 2024).

Equipment, materials or supplies

Across the included studies, a wide range of materials, resources and methods were used to deliver CBT Skills interventions. Most programmes appeared to be based on a structured curriculum or session plans. Some interventions, such as Daunic et al. (2006) and Schwartz-Mette et al. (2024), included fully scripted sessions to promote consistency.

Several programmes offered additional materials specifically for parents or teachers, including separate parent and teacher training resources described in Cole et al. (2013). To encourage parent participation, Lochman et al. (2017) reported practical supports such as dinner, a supervised childcare room, and a $5 attendance stipend. Both Lochman et al. (2017) and Orgilés et al. (2023) also reported web-based platforms. The Coping Power — Internet Enhanced (CP-IE) program (Lochman et al., 2017) offered tailored interfaces for different user groups (children, parents, teachers and counsellors), allowing progress tracking, access to video clips, worksheets, quizzes and the ability to earn points to trade for prizes. The website described in Orgilés et al. (2023) provided young people with online access to activities and weekly tasks. Meanwhile parents were provided with session summaries and objectives, with a therapist available via phone or email for guidance.

Some studies used screening measures to ensure participants met the inclusion criteria (Cole et al., 2013; Kettlewell & Kausch, 1983). In Esposito-Smythers et al. (2017), a recording device was required for audiotaped fidelity checks. The

programme described in Burraston et al. (2012), required mobile phones and supportive voice messages from significant others. Hence, technical infrastructure was needed to store these messages, automatically call the young person at scheduled times, and record the progress updates they provided in response.

Finally, activities were reported across studies for which learning materials would be required. These included games, stories and videos, prompts for group discussions, problem-solving tasks, breathing exercises and worksheets. For example, Avci and Kelleci (2016) reported video clips to illustrate concepts, in-class exercises, examples from newspapers, and homework assignments, while Glick and Goldstein (1987) reported use of a 'hassle log' where young people recorded anger provoking experiences. Studies which incorporated rewards utilised snacks, toys and funding for group activities.

Who delivers CBT Skills?

A wide range of professionals delivered the CBT Skills programmes, with many interventions involving input from multiple disciplines.

Most commonly, programmes were delivered by psychologists. These included qualified or trainee school psychologists, educational psychologists and clinical psychologists (Augustyniak et al., 2009; Cole et al., 2013; Ime, 2024; Johnsen et al., 2024; Kumuyi et al., 2022; Lesure-Lester, 2002; Njardvik et al., 2022; Schwartz-Mette et al., 2024; Squires & Caddick, 2012; Sukhodolsky et al., 2000; Te Brinke et al., 2021; Watson, 2009; Yeo & Choi, 2011).

School counsellors and graduate counselling students also frequently acted as programme facilitators (Augustyniak et al., 2009; Freiden, 2006; Lochman et al., 2017; McLaughlin, 2009; Wagner et al., 2014).

Beyond psychologists and counselling professionals, several interventions were delivered by teachers and school personnel (Cole et al., 2013; Daunic et al., 2006; Etscheidt, 1991; Moynahan & Strømgren, 2005; Squires & Caddick, 2012), therapists (Johnsen et al., 2024; Nwolisa, 2023; Poulin et al., 2001) and social workers (Hawkins et al., 1991; Logsdon, 2003; Te Brinke et al., 2021).

In seven studies, delivery was carried out or supported by researchers or student researchers (Avci & Kelleci, 2016; Bowman & Auerbach, 1982; Kozina, 2021; Logsdon, 2003; Oparaduru, 2017; Saba et al., 2023; Sukhodolsky et al., 2000).

Additional delivery roles included interventionists (Esposito-Smythers et al., 2017), coping power staff (Lochman et al., 2017), coaches (Pullen, 1996) and parent co-leaders (Poulin et al., 2001).

In Cook et al. (2014), the Becoming a Man (BAM) programme could be delivered by any college-educated facilitator, though in the included study it appeared to be delivered by psychology and social work graduates. Burraston et al. (2012) did not report the specific professionals delivering the CBT component, although participants could access professional counsellors in the year after the intervention.

Finally, in Orgilés et al. (2023), the intervention was delivered entirely through an online platform, supplemented by therapists who supported parents via phone or email and reviewed the children's submissions.

How was the intervention delivered?

Where reported, interventions were most commonly delivered in person, to groups or whole classes (Augustyniak et al., 2009; Avci & Kelleci, 2016; Burraston et al., 2012; Cole et al., 2013; Cook et al., 2014; Daunic et al., 2006; Esposito-Smythers et al., 2017; Etscheidt, 1991; Freiden, 2006; Glick & Goldstein, 1987; Hawkins et al., 1991; Ime, 2024; Karatas & Gokcakan, 2009; Kozina, 2021; Kumuyi et al., 2022; Logsdon, 2003; McLaughlin, 2009; Moynahan & Strømgren, 2005; Njardvik et al., 2022; Nwolisa, 2023; Oparaduru, 2017; Poulin et al., 2001; Pullen, 1996; Saba et al., 2023; Schwartz-Mette et al., 2024; Squires & Caddick, 2012; Sukhodolsky et al., 2000; Watson, 2009; Yeo & Choi, 2011). Two studies incorporated additional components beyond group delivery: in Burraston et al. (2012) the core CBT sessions were delivered in person, but the cell phone component was delivered individually and remotely; in Cook et al. (2014), the CBT component was delivered in groups, while the tutoring component was delivered face-to-face with pairs of students. Lesure-Lester (2002) reported that sessions were delivered to small groups of 2–3 students.

Some interventions were delivered entirely face-to-face with individuals (Bowman & Auerbach, 1982; Doğan & Çam, 2020; Te Brinke et al., 2021; Wagner et al., 2014),

while the intervention described by Kettlewell and Kausch (1983) used a mixed format, combining individual and group sessions. The programme in Johnsen et al. (2024) was delivered through joint sessions with the young person and their parents, alongside two sessions conducted solely with parents.

Two studies incorporated online or hybrid delivery formats. The Coping Power — Internet Enhanced (CP-IE) programme (Lochman et al., 2017) combined small-group, face-to-face sessions with online content accessed via a website. The entire intervention in Orgilés et al. (2023) was delivered remotely to individuals through an online platform.

Where CBT Skills is delivered

CBT Skills programmes were most frequently delivered in school settings (Augustyniak et al., 2009; Avci & Kelleci, 2016; Cole et al., 2013; Cook et al., 2014; Daunic et al., 2006; Doğan & Çam, 2020; Etscheidt, 1991; Freiden, 2006; Ime, 2024; Johnsen et al., 2024; Karatas & Gokcakan, 2009; Kozina, 2021; Kumuyi et al., 2022; Lochman et al., 2017; Logsdon, 2003; McLaughlin, 2009; Moynahan & Strømgren, 2005; Njardvik et al., 2022; Saba et al., 2023; Schwartz-Mette et al., 2024; Squires & Caddick, 2012; Sukhodolsky et al., 2000; Te Brinke et al., 2021; Wagner et al., 2014; Watson, 2009; Yeo & Choi, 2011). Whilst the child component of Lochman et al.'s (2017) study was delivered in schools, the parent component was delivered in small groups within the community. One additional programme was delivered in a learning centre (Bowman & Auerbach, 1982).

Two interventions were delivered remotely via a website: Orgilés et al. (2023) was entirely online, and Lochman et al. (2017) was partially online.

Other programmes were provided in a range of community and specialist settings, including:

- A community setting (Esposito-Smythers et al., 2017)
- A summer day camp affiliated with a residential treatment centre (Kettlewell & Kausch, 1983)
- Juvenile justice correctional facilities (Glick & Goldstein, 1987; Hawkins et al., 1991; Nwolisa, 2023; Pullen, 1996)

In Burraston et al. (2012), the setting for the CBT programme was not reported, although the mobile phone component was delivered remotely by phone.

Training for the providers of CBT Skills

Where stated, the level and type of training provided to programme facilitators varied considerably. In many interventions, formal training sessions were delivered prior to programme implementation (Augustyniak et al., 2009; Daunic et al., 2006; Esposito-Smythers et al., 2017; Etscheidt, 1991; Ime, 2024; Johnsen et al., 2024; Lochman et al., 2017; Logsdon, 2003; Moynahan & Strømgren, 2005; Njardvik et al., 2022; Pullen, 1996; Schwartz-Mette et al., 2024; Squires & Caddick, 2012; Sukhodolsky et al., 2000; Te Brinke et al., 2021; Wagner et al., 2014).

Some studies provided further detail about this training, for example Daunic et al. (2006) noted that training covered both the rationale for the intervention and the session content, while Johnsen et al. (2024) described an intensive model consisting of a six-day training course followed by a booster session.

Supervision and ongoing support were also common features of programme delivery. Several studies reported weekly or regular supervision, such as weekly group supervision in Johnsen et al. (2024), weekly individual supervision in Lochman et al. (2017), and a combination of monthly review meetings with weekly telephone supervision in Logsdon (2003). Ongoing weekly support was also noted in Njardvik et al. (2022), Schwartz-Mette et al. (2024) and Sukhodolsky et al. (2000), while 24 hours of supervision was reported in Moynahan and Strømgren (2005). In Etscheidt (1991), facilitators met weekly to review the previous session and prepare for the next.

Some studies incorporated fidelity monitoring and other feedback mechanisms. Njardvik et al. (2022) reported that random sessions were reviewed and feedback provided, while Esposito-Smythers et al. (2017) noted audiotaped fidelity checks occurred alongside supervised practice.

In other programmes, training arrangements differed. Cole et al. (2013) described basic training for class teachers and parents, but did not specify any training provided to the trainee educational psychologists who delivered the intervention. Training was not required for delivering the Becoming a Man programme, although Cook et al. (2014) noted that facilitators were required to be college graduates. In Kumuyi et al. (2022), facilitators were licensed clinicians who had completed supervised psychotherapy internships using treatment manuals. In Freiden (2006), no formal training was reported, however the facilitator developed

the intervention. Hawkins et al. (1991) did not report on training but noted that facilitators had prior CBT experience.

Duration and intensity of CBT Skills

Where intensity was reported, sessions were most frequently delivered on a weekly basis (Augustyniak et al., 2009; Avci & Kelleci, 2016; Burraston et al., 2012; Cole et al., 2013; Cook et al., 2014; Doğan & Çam, 2020; Esposito-Smythers et al., 2017; Ime, 2024; Karatas & Gokcakan, 2009; Kumuyi et al., 2022; Logsdon, 2003; McLaughlin, 2009; Njardvik et al., 2022; Nwolisa, 2023; Oparaduru, 2017; Poulin et al., 2001; Schwartz-Mette et al., 2024; Squires & Caddick, 2012; Sukhodolsky et al., 2000; Watson, 2009). Some studies included additional components beyond the weekly sessions: Burraston et al. (2012) incorporated a mobile phone component with twice daily contact for a year after the programme; and Cook et al. (2014) provided a daily tutoring component.

A smaller number of studies reported sessions every other week (Lesure-Lester, 2002; Lochman et al., 2017; Pullen, 1996), though for Lochman et al. (2017) it is likely that remote content was delivered more frequently. In Saba et al. (2023), sessions occurred once or twice weekly.

Four studies reported sessions delivered twice a week (Freiden, 2006; Hawkins et al., 1991; Kettlewell & Kausch, 1983; Orgilés et al., 2023), while a further two studies reported delivery two to three times per week (Bowman & Auerbach, 1982; Daunic et al., 2006).

Finally, three studies reported delivery three to four times a week (Etscheidt, 1991; Glick & Goldstein, 1987; Wagner et al., 2014).

The duration of CBT Skills interventions varied substantially across studies.

Some programmes were relatively brief, lasting two to three weeks (Esposito-Smythers et al., 2017; Etscheidt, 1991) or a month (Bowman & Auerbach, 1982; Freiden, 2006; Kettlewell & Kausch, 1983; McLaughlin, 2009; Orgilés et al., 2023).

Most interventions, however, lasted between one and three months (Augustyniak et al., 2009; Avci & Kelleci, 2016; Burraston et al., 2012; Cole et al., 2013; Daunic et al., 2006; Doğan & Çam, 2020; Glick & Goldstein, 1987; Hawkins et al., 1991; Ime, 2024; Karatas & Gokcakan, 2009; Kozina, 2021; Kumuyi et al., 2022; Logsdon, 2003; Moynahan & Strømgren, 2005; Njardvik et al., 2022; Nwolisa, 2023; Oparaduru, 2017; Poulin et al., 2001; Saba et al., 2023; Schwartz-Mette et al., 2024; Squires & Caddick,

2012; Sukhodolsky et al., 2000; Watson, 2009; Yeo & Choi, 2011). Some studies included additional components beyond the core intervention period: in Burraston et al. (2012), the mobile phone component provided an additional year of support; and in Daunic et al. (2006), six booster lessons were delivered over the second half of the school year.

A smaller number of interventions were notably longer, lasting three to six months (Johnsen et al., 2024; Lesure-Lester, 2002; Lochman et al., 2017; Pullen, 1996). One programme (Cook et al., 2014) ran for an entire school year.

How Effective is the Intervention?

This section examines the effectiveness of CBT Skills in reducing violence, crime and offending, and other related outcomes through a systematic review and meta-analysis. It provides a robust and objective summary of existing evidence, incorporating advanced statistical techniques, including robust variance estimators (Pustejovsky & Tipton, 2022), for improved accuracy.

Quantitative data from **37 effectiveness studies,** including 2 papers that contributed independent clusters from multi-arm trials (Etscheidt, 1991; McLaughlin, 2009) provided information across a variety of **251 outcomes** related to the impact of CBT Skills for children and young people.

These studies, included in the meta-analysis, assessed the effectiveness of CBT Skills across a total of **4,154 children and young people.** These 37 effectiveness studies employed a range of study designs, including:

- Quasi-Experimental Designs (QED): (n =4, 11%)
- Randomised Controlled Trials (RCT): (n = 33, 89%)

The 37 effectiveness studies varied considerably in their methodological design and reporting characteristics (see Appendix 3 for individual study details).

All studies used a prospective design and assigned participants at the start of the study. Most studies allocated at the level of individual participants (n=36, 97%), while Watson (2009) allocated whole clusters (n=1; 3%)

The effectiveness studies have spanned the past 5 decades, with the earliest conducted by Barrett et al. (1982) and the most recent by Ime (2025).

The studies were conducted in 11 different countries including:

- USA: (n = 21, 57%)

- Turkey (n=4; 11%)
- Nigeria: (n = 3, 8%)
- England: (n = 2, 5%)
- Norway: (n = 1, 3%)
- Netherlands: (n = 1, 3%)
- Denmark: (n = 1, 3%)
- Iceland: (n = 1, 3%)
- Pakistan: (n = 1, 3%)
- Singapore (n = 1, 3%)
- Spain: (n = 1, 3%)

Fuller descriptions of the included studies are available in Appendix 3. Studies were assessed for methodological quality using the YEF-EQA critical appraisal tool and were rated as follows:

- High: (n = 4, 11%)
- Moderate: (n = 24, 65%)
- Low: (n = 9, 24%)

24 studies did not report their funding source (65%). Where reported across the remaining 13 studies, funding mainly came from Government and National Institutes (n=11; 30%), while two were funded by Foundations, Councils, and Academic Institutions (n=2; 5%). Many of the studies explicitly declared they had no conflict of interest (n = 11; 30%). Two of the included studies were flagged by the reviewer as having a conflict of interest (5%); Freiden (2006) and LeSure-Lester (2002) both developed and implemented the programme they evaluated. The remaining 24 studies did not declare or include a conflict-of-interest statement (65%).

The intensity of interventions varied from daily (n=2; 5%) to 1–2 times per week (n=19; 51%), 3–4 times per week (n=11; 30%), once a week (n=3; 8%), and every two weeks (n=2; 5%).

Interventions also varied in duration. Ten studies (27%) reported a total duration of between 1 week and 1 month, 20 studies (54%) reported a duration of 1 to 3 months, 6 studies (16%) reported a duration of 3 to 6 months, and 1 study (3%) reported a duration of 6 months to 1 year.

Given the significant heterogeneity in the time commitment, frequency, and duration of the interventions, treatment intensity was classified into three distinct tiers to facilitate the moderator analysis for these CBT Skills programmes. High-intensity programmes (n=4; 11%) involved massive time commitments spanning dozens of sessions, an entire academic year, or multiple years, and often utilised a comprehensive, multi-systemic approach (e.g., combining universal classroom instruction, targeted groups, and family therapy, or integrating intense behavioural therapy with daily academic remediation). Moderate-intensity programmes (n=21; 57%) represented a standard clinical or school-based approach, typically spanning 10 to 20 sessions (roughly 12 to 15 hours of contact time) delivered over a few months, and occasionally integrated secondary components like booster sessions or brief parent involvement. Low-intensity interventions (n=12; 32%) were highly targeted and brief, ranging from just a few hours in total to around 5 to 8 sessions. Further details on programme intensity are provided in Appendix 6.

Interventions were predominantly delivered in school or educational settings (n = 25, 68%). Other settings included justice, custodial and community settings (n = 4, 11%), residential or social care settings (n = 3, 8%), or community based clinical and mental health services (n = 3, 8%). Two studies delivered interventions using a digital computerised format (5%).

Most interventions were delivered either by mental health or therapeutic professionals (n=15, 41%), or by the study researchers and their teams (n=14, 38%). A further five studies (13%) reported delivery by school staff, while three studies (8%) did not specify who delivered the intervention.

Demographic characteristics

Study inclusion criteria varied across the 37 studies. Most studies included children and young people with externalising or behavioural dysregulation difficulties (n=21, 57%). Smaller numbers of studies focused on justice-involved or offending-related children and young people (n=5, 14%), those with mental health

or emotional difficulties (n=3, 8%), those with school-risk or disengagement profiles (n=3, 8%), and universal or whole-population school samples (n=3, 8%). A further two studies (5%) included children and young people with social care involvement or exposure to adversity.

In terms of gender composition, the included studies were predominantly male. Eighteen studies (49%) included mixed-sex samples, 12 studies (32%) included majority-male samples, and 7 studies (19%) included male-only samples. No studies included majority-female or female-only samples.

The included studies covered a range of developmental stages. Of the 37 studies, 11 (30%) focused on childhood, preadolescence, and early adolescence (ages 5–11), 8 (22%) on early adolescence (ages 12–15), and 5 (14%) on middle to late adolescence (ages 16 and over). In addition, 13 studies (35%) included mixed age groups spanning more than one developmental stage.

Reporting on the ethnic and racial composition of participants varied across the literature. Of the 37 included studies, 16 (43%) did not report any ethnicity-related demographic information. The remaining 21 studies provided specific statistical breakdowns of participants' race and/or ethnicity. Among these, four studies (11%) included samples composed entirely of young people from Black and Global Majority backgrounds. A further 10 studies (27%) featured highly diverse samples in which young people from Black and Global Majority backgrounds constituted the majority. Three studies (8%) reported mixed samples that were majority white, but still included a substantial proportion of young people from Black and Global Majority backgrounds, representing between 15% and 49% of participants. The remaining four studies (11%) reported predominantly white samples.

Reporting on socio-economic status was similarly inconsistent. Nine studies (24%) indicated that the majority of families in the sample were on low incomes, five (14%) featured predominantly low- to middle-income families, and three (8%) reported mixed income distributions. The remaining 20 studies (54%) did not provide any information on participants' socio-economic status.

Finally, only two studies reported information relevant to care experience. These studies suggest that some participants had experienced substantial family instability and adversity. One study reported that 43% of participants were recruited following discharge to home from a youth shelter, where they had been

placed temporarily because of volatile and unstable home environments (Esposito-Smythers, 2017). Another study described participants as highly vulnerable, having experienced significant psychosocial trauma, including maternal deprivation, parental abandonment, family breakdown, physical abuse, and the harsh realities of surviving on the streets and within overcrowded correctional facilities (Nwolisa, 2023).

Measured outcomes

Across the 37 effectiveness studies, 251 outcomes were identified and extracted, spanning **17 outcome categories** within the YEF Outcomes Framework:[1]

- Ability to resolve conflicts: (k=6, n=4)
- Behavioural difficulties: (k=68, n=27)
- Building and maintaining relationships: (k=9, n=4)
- Bullying: (k=7, n=4)
- Crime and offending: (k=11, n=4)
- Criminal peers: (k=6, n=1)
- Drug and alcohol use: (k=18, n=4)
- Family Relationships and Support: (k=1, n=1)
- General Mental Health: (k=26, n=7)
- Helping others: (k=5, n=4)
- Meaningful Relationships: (k=3, n=2)
- Opportunities for Education, Employment and Training: (k=1, n=1)
- Parenting practices: (k=2, n=1)
- Regulating and Managing Emotions: (k=54, n=15)
- School Engagement: (k=16, n=5)

[1] The YEF Outcomes Framework identifies specific outcomes linked to reducing the risk of children and young people becoming involved in crime and violence, providing a structured approach for measuring the impact of interventions

- Self-esteem: (k=16, n=6)
- Victim of crime: (k=2, n=2)

To inform our violence estimate, outcomes must be a measure of violent crime or violent offending, or a measure of overt or relational aggression that was directed at another person. Measures of bullying can inform the violence estimate where the study reports bullying as repeated physical violence or repeated verbal harassment.

Across the 22 studies that met this criteria, we identified 56 violence outcomes. Thirty-eight were also coded to behavioural difficulties (68%), 6 related to crime and offending (11%), 7 coded to bullying (13%) and 5 measures of outward anger were also coded to regulating and managing emotions (9%).

Collectively, these measures assess aggression, including physical aggression, verbal aggression, bullying, conduct problems and general behavioural difficulties. Additionally, proneness to maladaptive behaviour was also used as a proxy in one study to measure violent crimes (e.g., rape). These measures aid in capturing various forms of aggressive and offending behaviour across youth populations.

Outcomes were derived in several ways, most predominantly through a multi-informant approach utilising standardised survey measures (k=219; 87%), Clinician assessment/ observation (k=16; 6%), and administrative/school records (k=16; 6%).

Survey measures were reported by parents (k=21; 8%), peers (k=2; 0.8%), and teachers (k=38; 15%), but most self-reported survey measures were completed by the young people themselves (k=158; 63%).

We present summary results from two separate multivariate meta-analyses on violence, and crime and offending below (Table 1).

Table 1: *Summary of findings on violence and crime outcomes*

Outcome	SMD	CI (95%)	*P*	% reduction	Impact rating	Number of studies	Evidence rating
Violence only	-0.77	-1.40, -0.13	0.019*	72%	High	22	3

Crime and Offending	-0.79	-0.93, -0.65	<.0001***	71%	High	4	1

Meta-analysis of violence outcomes in CBT Skills interventions

CBT Skills is estimated to have a high impact on violence, corresponding with a 72% reduction in violence, based on 56 measured outcomes across 22 studies.

The primary focus in the initial analysis is the reduction and prevention of violence, as defined by YEF. Violence is a broad construct that incorporates incidents/behaviours as well as convictable offences. Violence may be of a physical, verbal, psychological, or sexual nature (YEF, 2023: p.12).

The team identified 56 outcomes measuring **violence** specifically across 22 studies: 3 Type A, 16 Type C, and 3 Type D, giving an overall security rating of **Level 3**.

A total of $k = 56$ outcomes were included in the analysis. The estimated average outcome based on the random-effects model was $\hat{\mu} = -0.77$, (95% CI: -1.41 to -0.13); p<0.05 (Table 2).

This estimate was statistically significantly different from zero $z = -2.35$, $p < 0.05$, and the effect remained the same (g=-0.77) when adjusting for clustering across studies using robust variance estimation, the p-value increased slightly, but remained significant (t = -2.36, df = 20.7, p = 0.028).

***Table 2**: RVE Output for meta-analysis on violence outcomes*

	Estimate	SE	t-stat	d.f (Satt)	p-val (Satt)	Sig	YEF impact rating	YEF evidence security rating
Intercept	-0.77	0.33	-2.36	20.7	0.028	*	High	3

In practical terms, this indicates that the intervention had a meaningful impact on violence-related outcomes across 22 studies. Based on YEF impact

categorisation, the effect size (g = -0.77) corresponds to a "high impact" rating. As the p-value is below .05, the reader can have some confidence in these findings. The prediction interval is very wide (-3.74 to 2.07), and gives us the range of effects a new (but similar) study might plausibly find.

According to the Q-test, high heterogeneity was present; I^2 = 95.6%, Q(df = 55) = 145.895, p =1236.856. This level of variability suggests that the studies are inconsistent with each other, beyond what would be expected.

The SMD of -0.77 corresponds to a relative risk reduction of 71.9% in the risk of re-committing violence, with an absolute risk reduction of 11.5%. Among those who participate in CBT skills, 5% go on to be involved in violence, compared to 16% of those who do not receive CBT skills (Figure 1).

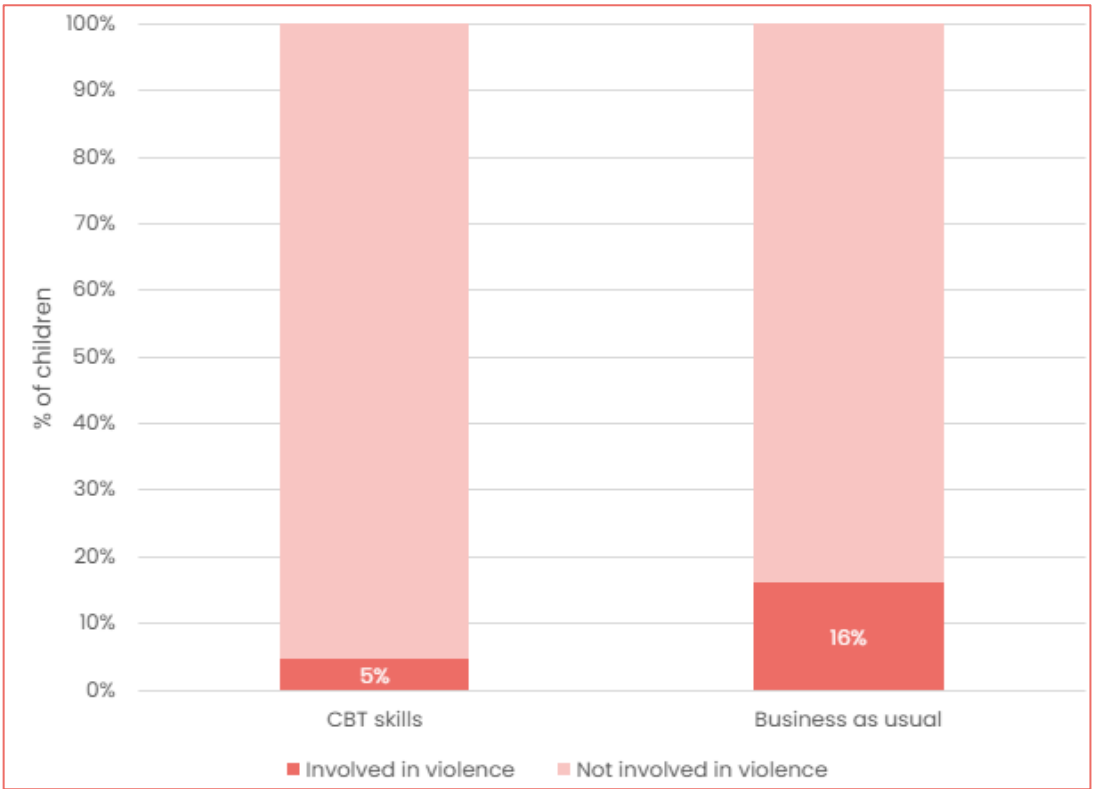

Figure 1: *Absolute risk reduction in violence from CBT skills interventions compared to business as usual*

Meta-analysis of crime and offending outcomes in CBT Skills interventions

CBT skills may have a high impact on crime and offending, corresponding with a 71% decrease in crime and offending, based on 11 measured outcomes across 4 studies. As this is based on very low-security evidence (Level 1), this finding should be treated with substantial caution.

The team identified 11 outcomes measuring **crime and offending** specifically across four studies: 2 Type C, and 2 Type D giving an overall security rating of **Level 1**. Most Crime and offending outcomes measured official criminal involvement and recidivism e.g. 'general recidivism' and 'total convictions' with several outcomes measuring specific adverse legal consequences such as 'arrests'.

Most crime and offending outcomes looked at 'proneness to maladaptive behaviour', which includes stealing and rape. Several outcomes measure official criminal involvement and recidivism e.g., 'new crime', 'time to rearrest' and 'never rearrested'. Other outcomes looked at 'technical violations' e.g., breaking probational terms.

A total of $k = 11$ outcomes were included in the analysis. The estimated average outcome based on the multilevel random-effects model was $\hat{\mu} = -0.79$, (95% CI: --0.93, -0.65). This estimate was statistically significantly different from zero in the initial model z = -10.91, p <0001. When adjusting for clustering across studies using robust variance estimation, the effect remained the same (g = -0.79), but although the p-value increased slightly it remained statistically significant at the 0.05 level (t = -9.25, df = 1.37, p = 0.0338) (Table 3).

Table 3: *RVE Output for meta-analysis on crime and offending outcomes*

	Estimate	SE	t-stat	d.f (Satt)	p-val (Satt)	Sig	YEF impact rating	YEF evidence security rating
Intercept	-0.79	0.09	-9.25	1.37	0.0338	*	High	1

In practical terms, this indicates that the intervention had a large and meaningful impact on crime-related outcomes. Based on YEF impact categorisation, the effect size (g = -0.79) corresponds to a high impact rating. The prediction interval is quite narrow (-0.99 -0.59), giving the range of effects that a new but similar study might plausibly find. This suggests that the average effect is strongly beneficial, and future studies are likely to find similar favourable effects.

According to the Q-test, high heterogeneity was present; I^2 = 68.8%, Q(df = 10) = 32.094, p < 0.001. This level of variability suggests inconsistency between studies.

The SMD of -0.79 corresponds to a relative risk reduction of 70.5% in the risk of being involved in crime and offending, with an absolute risk reduction of 17.6%. In simple terms, among those who participate in CBT skills, 7% go on to be involved in crime or offending, compared to 25% of those who do not receive CBT skills (Figure 2).

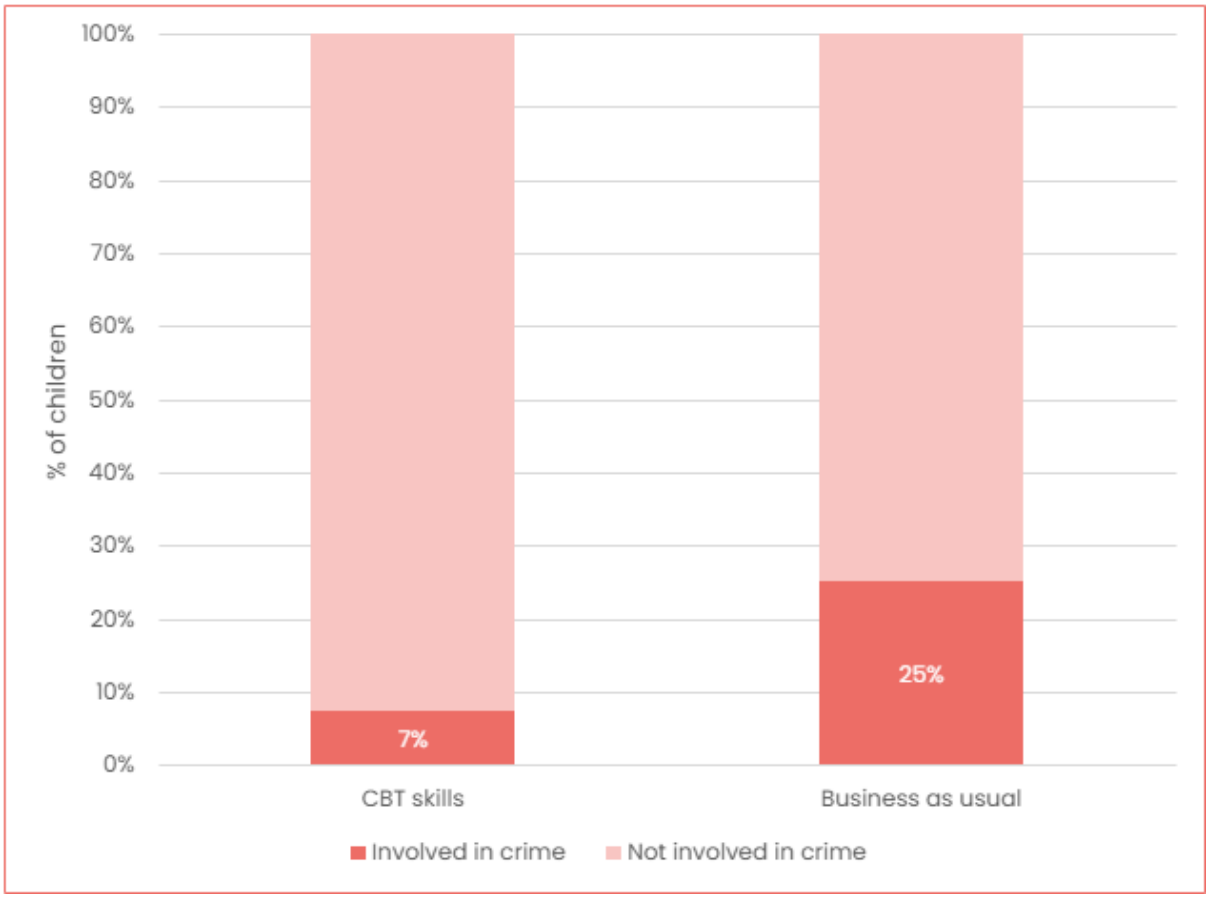

Figure 2: *Absolute risk reduction in crime and offending from CBT Skills interventions compared to business as usual*

Meta-analysis of all related outcomes in CBT Skills interventions

Including the violence, crime, and offending outcomes described above, the team extracted a total of 251 measured outcomes from 37 effectiveness studies that align with the YEF outcomes framework. These studies included 4 Type A, 24 Type C, 9 Type D, giving an overall security rating of **Level 3**.

A total of $k = 251$ outcomes were included in the analysis. The estimated average outcome based on the random-effects model was $\hat{\mu} = -0.58$, (95% CI: -0.92, -0.24).

This estimate was statistically significantly different from zero, z = -3.33, p <0.001, and remained statistically significant when adjusting for clustering across studies using robust variance estimation (t = -3.33, df = 35.5, p <0.005). The estimate remained the same (g = -0.58) suggesting the original multivariate meta-analysis was stable.

Table 4: *RVE Output for meta-analysis of all related outcomes*

	Estimate	SE	t-stat	d.f (Satt)	p-val (Satt)	Sig	YEF impact rating	YEF evidence security rating
Intercept	-0.58	0.17	-3.33	35.5	<0.005	**	High	3

In practical terms, this indicates that CBT Skills had a meaningful overall impact across the full set of outcomes examined. The pooled effect size suggests that, on average, participants receiving CBT Skills had better outcomes than those in the comparison conditions. However, the prediction interval was quite wide (-2.63 to 1.47), giving the range of effects that a new but similar study might plausibly find. This suggests that although the average effect is beneficial, future studies may still find smaller effects or no clear effect.

According to the Q-test, high heterogeneity was present; I^2 = 96.3%, Q(df = 250) = 6772.056, p <.001. This level of variability suggests substantial inconsistency across studies, beyond what would be expected by chance alone.

Such significant variability suggests that the differences in effect sizes are not solely due to random chance but may be influenced by specific study characteristics or contexts. To explore potential sources of this heterogeneity, conducting moderator analyses is recommended.

Subgroup analysis

Details on the moderator categories are provided in Appendix 6.

Intensity of the programme

To explore whether programme effectiveness varied by intensity of the programme, subgroup meta-analyses were conducted (Table 5).

Table 5: *Subgroup analysis on programme intensity*

Intensity of intervention	k	n	SMD (robust SE)	p-value	YEF impact rating
High intensity / Comprehensive & Multi-Systemic Interventions	18	4	−0.36 (0.11)	.085	High
Moderate intensity / Standard Interventions	155	21	−0.70 (0.23)	.011 *	High
Low intensity / Brief Interventions	78	12	−1.09 (0.40)	.032 *	High
Between-group heterogeneity (Q_n test); 3 subgroups; Q_n = 2.27; p = .104					

Note. Estimates computed using robust variance estimation (RVE) with CR2 adjustment for dependent effect sizes (Tipton, 2015). Subgroup-specific p-values are based on robust t-tests with Satterthwaite small-sample degrees of freedom. The omnibus test of differences across outcome domains was conducted using a CR2-robust Wald chi-square test. Significance levels: p < .05 (*), p < .01 (**), p < .001 (***).

Subgroup analysis by intervention intensity suggested that CBT Skills interventions were associated with beneficial effects across all three intensity categories, although the magnitude of effect and statistical significance varied. Low-intensity or brief interventions showed the largest effect size and reached statistical significance (SMD = −1.09, k = 78, n = 12, p = .032), suggesting that minimal exposure to CBT skills training may be sufficient to produce substantial improvements. Moderate-intensity or standard interventions demonstrated a smaller, but statistically significant beneficial effect (SMD = −0.70, k = 155, n = 21, p = .011). High-intensity or comprehensive interventions showed a smaller effect size (SMD = −0.36, k = 18, n = 4), which did not reach conventional levels of statistical significance (p = .085)Although the point estimates appear to suggest stronger effects for low-intensity interventions than for moderate- or high-intensity interventions, the formal test of between-group heterogeneity was not statistically significant (Q_n = 2.27, p = .104). This indicates that there is insufficient evidence to conclude that intervention effects differ reliably according to intensity. In practical terms, CBT Skills interventions appear beneficial across intensity levels, but the present analysis does not support intensity as a clear moderator of effectiveness.

Inclusion criteria of the participants

To explore whether programme effectiveness varied based on the inclusion criteria of the trial, subgroup meta-analyses were conducted (Table 6).

Table 6: *Subgroup analysis on inclusion criteria*

Participant inclusion criteria	k	n	SMD (robust SE)	p-value	YEF impact rating
Externalising / behavioural dysregulation	151	21	−0.78 (0.22)	.005 **	High
School-risk / disengagement profile	29	3	−0.13 (0.14)	.459	Moderate

Mental health / emotional difficulties	28	3	−1.57 (1.04)	.312	High
Justice-involved / offending-related CYP	23	5	−1.38 (0.49)	.089 †	High
Universal / whole-population school sample	13	3	−0.12 (0.15)	.533	Moderate
Social care / adversity-exposed CYP	7	2	−0.59 (0.28)	.280	High
Between-group heterogeneity (Q_n test); 6 categories; Q_n = 3.04; p<.009 **					

CBT Skills interventions showed beneficial effects across all participant inclusion criteria, with all subgroup estimates in the negative direction, indicating better outcomes for participants receiving CBT Skills interventions than for controls. The clearest evidence was observed for samples selected on the basis of externalising / behavioural dysregulation (k = 151 effect sizes, 21 studies), which showed a statistically significant beneficial effect (SMD = −0.78, p = .005). This was also by far the largest evidence base, making it the most robust finding within this subgroup analysis.

A relatively large beneficial effect was also observed for justice-involved / offending-related CYP (SMD = −1.38), although this was only marginally significant (p = .089). Similarly, mental health / emotional difficulties showed the largest point estimate (SMD = −1.57), but this subgroup was based on only three studies and had a large standard error, so the estimate was imprecise and not statistically significant (p = .312). Smaller, non-significant beneficial effects were observed for social care / adversity-exposed CYP (SMD = −0.59), school-risk / disengagement profile (SMD = −0.13), and universal / whole-population school samples (SMD = −0.12).

Finally, a Q-test for between-group heterogeneity indicated that the effects of CBT Skills interventions differed significantly across participant inclusion criteria. A robust Wald chi-square test showed significant between-group heterogeneity (χ^2 (5) = 3.04, p = .009), providing evidence that intervention effects vary depending on the type of young people included in the sample. In practical terms, this suggests that CBT Skills interventions may be particularly effective for children and young people with externalising or behavioural dysregulation, while evidence for other inclusion profiles remains more limited or uncertain.

Intervention Setting

To explore whether programme effectiveness varied by intervention setting, subgroup meta-analyses were conducted (Table 7).

Table 7: *Subgroup analysis on intervention setting*

Intervention setting	k	n	SMD (robust SE)	p-value	YEF impact rating
School / education setting	170	26	−0.73 (0.20)	.003 **	High
Community-based clinical / mental health service	30	3	−0.11 (0.14)	.509	Moderate
Justice / custodial / community supervision setting	22	4	−1.28 (0.61)	.170	High
Computerised / digital	15	2	−2.38 (0.67)	.174	High
Residential care / social care setting	14	3	−0.59 (0.30)	.218	High

Between-group heterogeneity (Q_n test); 5 settings; Q_n = 4.39; p<.002 **

CBT Skills interventions showed beneficial effects across all intervention settings, with all subgroup estimates in the negative direction, indicating better outcomes for participants receiving CBT Skills interventions than for controls. The clearest evidence was observed in school / education settings (k = 170 effect sizes, 26 studies), where CBT Skills interventions produced a statistically significant beneficial effect (SMD = −0.73, p = .003). This was also the subgroup with by far the largest evidence base.

The remaining settings all showed negative point estimates, suggesting beneficial effects in the same direction, but none reached conventional levels of statistical significance. In particular, the estimated effects were particularly large for computerised / digital delivery (SMD = −2.38) and for justice / custodial / community supervision settings (SMD = −1.28), but these estimates were based on very small numbers of studies and had low Satterthwaite degrees of freedom, so they should be interpreted cautiously. More modest, non-significant beneficial effects were also observed in residential care / social care settings (SMD = −0.59) and community-based clinical / mental health services (SMD = −0.11).

Finally, a Q-test for between-group heterogeneity indicated that the effects of CBT Skills interventions did differ significantly across intervention settings. A robust Wald chi-square test showed significant between-group heterogeneity ($\chi^2(4)$ = 4.39, p = .002), providing evidence that intervention setting helps explain variation in effect sizes. In practical terms, this suggests that CBT Skills interventions may be more effective in some contexts than others, with the strongest and most reliable evidence currently coming from school / education settings.

Intervention delivery format

To explore whether programme effectiveness varied by mode of delivery, subgroup meta-analyses were conducted (Table 8).

Table 8: *Subgroup analysis on mode of delivery*

Delivery format	k	n	SMD (robust SE)	p-value	
Large groups (15+), face to face	78	8	−0.98 (0.33)	.049 *	High
Medium groups (6–14), face to face	42	8	−0.42 (0.36)	.309	High
Individual, face to face	41	5	−0.36 (0.16)	.099 †	High
Groups (size unclear), face to face	35	7	−0.95 (0.51)	.137	High
Small groups (2–5), face to face	29	5	−0.47 (0.39)	.333	High
Hybrid interventions (Face-to-Face + Technology/Media)	18	3	−2.18 (0.85)	.188	High
Individuals and small groups (5 per group), face to face	8	1	−0.33 (0.00)	.007 **	High
Between-group heterogeneity (Q_n test); 7 formats; Q_n = 1.71; p=.113					

CBT Skills interventions showed beneficial effects across all delivery formats, with all subgroup estimates in the negative direction, indicating better outcomes for participants receiving CBT Skills interventions than for controls. The strongest evidence, based on the number of effect sizes and studies, was observed for large groups (15+) delivered face to face (k = 78 effect sizes, 8 studies), which showed a statistically significant beneficial effect (SMD = −0.98, p = .049). Individual, face-

to-face delivery also showed a beneficial effect of smaller magnitude (SMD = −0.36), although this was only marginally significant (p = .099).

A statistically significant effect was also found for interventions delivered to individuals and small groups of 5 (SMD = −0.33, p = .007), but this subgroup was based on only one study and should therefore be interpreted with substantial caution. Other formats, including medium groups, small groups, groups of unclear size, and hybrid interventions, all showed beneficial point estimates, but none reached conventional levels of statistical significance.

Finally, a Q-test for between-group heterogeneity indicated that the effects of CBT Skills interventions did not differ significantly across delivery formats. A robust Wald chi-square test found no clear evidence of between-group heterogeneity ($\chi^2(6) = 1.71$, p = .113), suggesting that intervention effects were broadly similar across formats overall. In practical terms, this means there is no strong statistical evidence that one delivery format was more effective than another, even though large-group face-to-face delivery appeared the most promising in the subgroup-specific estimates.

Family/parent Involvement

To explore whether programme effectiveness varied due to the additive effects of combining child CBT with family therapy or Parent Management Training, subgroup meta-analyses were conducted (Table 9).

Table 9: *Subgroup analysis on family involvement*

Family involvement	k	n	SMD (robust SE)	p-value	YEF impact rating
No family component	180	27	−0.82 (0.19)	<.001 ***	High
Informational support and homework	39	6	−1.23 (0.70)	.156	High

facilitation (low involvement)					
Integrated joint family sessions (high involvement)	22	2	−0.25 (0.02)	.062 †	Moderate
Parallel parent management training (moderate involvement)	10	2	+0.13 (0.19)	.604	Harmful
Between-group heterogeneity (Q_n test); 3 groups; Q_n = 0.23; p = 0.795					

CBT Skills interventions showed varying effects depending on the degree of family involvement. The clearest evidence was for interventions with no family component (k = 180 effect sizes, 27 studies), which showed a statistically significant beneficial effect (SMD = −0.82, p < .001). This subgroup also had by far the largest evidence base, suggesting that CBT Skills interventions can be effective even when delivered without a family component.

Interventions involving informational support and homework facilitation also showed a large beneficial point estimate (SMD = −1.23), but this effect was not statistically significant (p = .156). Integrated joint family sessions showed a smaller beneficial effect (SMD = −0.25) that was marginally significant (p = .062), although this estimate was based on only two studies and should therefore be interpreted cautiously. In contrast, parallel parent management training was the only subgroup with a positive point estimate (SMD = +0.13, p = .604), suggesting no evidence of benefit in this category.

Finally, a Q-test for between-group heterogeneity indicated that the effects of CBT Skills interventions differed significantly across family involvement categories. A robust Wald chi-square test showed significant between-group heterogeneity ($\chi^2(3) = 4.92$, p = .002), providing evidence that the impact of CBT Skills

interventions varies according to how families are involved. In practical terms, this suggests that family involvement may shape intervention effectiveness, although the strongest and most reliable evidence in this analysis came from interventions with no family component.

Outcome type

To explore whether programme effectiveness varied due to the outcome measured, subgroup meta-analyses were conducted (Table 10).

Table 10: *Subgroup analysis on outcome domain*

Outcome Domain	k	n	SMD (robust SE)	p-value	YEF impact rating
Behavioural difficulties	68	27	−0.52 (0.17)	.005 **	High
Regulating and managing emotions	54	15	−1.01 (0.36)	.025 *	High
General mental health	26	7	−1.60 (0.59)	.052 †	High
Drug and alcohol use	18	4	−0.17 (0.14)	.309	Moderate
School engagement	16	5	−0.28 (0.05)	.029 *	High
Self-esteem	16	6	−1.08 (0.41)	.064 †	High
Crime and offending	11	4	−0.68 (0.11)	.026 *	High

Building and maintaining relationships	9	4	−0.99 (0.68)	.252	High
Bullying	7	4	−1.57 (0.83)	.163	High
Ability to resolve conflicts	6	4	−0.27 (0.68)	.718	High
Criminal peers	6	1	−0.67 (0.00)	n/a †	
Helping others	5	4	−0.44 (0.12)	.043 *	High
Meaningful relationships	3	2	−1.84 (1.16)	.360	High
Parenting practices	2	1	−0.26 (0.00)	n/a †	
Victim of crime	2	2	−0.51 (0.03)	.031 *	High
Family Relationships and Support	1	1	−0.63 (0.00)	n/a †	

CBT Skills interventions demonstrated beneficial effects across almost all outcome domains, with nearly all subgroup estimates in the negative direction, indicating better outcomes for participants receiving CBT Skills interventions than for controls. The strongest evidence came from behavioural difficulties (k = 68 effect sizes, 27 studies), where CBT Skills interventions showed a statistically significant beneficial effect (SMD = −0.52, p = .005). Significant beneficial effects were also found for regulating and managing emotions (SMD = −1.01, p = .025), school engagement (SMD = −0.28, p = .029), crime and offending (SMD = −0.68, p

= .026), helping others (SMD = −0.44, p = .043), and victim of crime (SMD = −0.51, p = .031).

Some additional domains showed promising but less certain evidence. General mental health (SMD = −1.60, p = .052) and self-esteem (SMD = −1.08, p = .064) were both marginally significant, suggesting potentially meaningful benefits but with substantial uncertainty. Other domains, including building and maintaining relationships, bullying, drug and alcohol use, meaningful relationships, and ability to resolve conflicts, all had beneficial point estimates but did not reach conventional levels of statistical significance.

A small number of very sparse domains produced apparently highly significant estimates, including criminal peers and parenting practices, but these results were based on single studies and had near-zero estimated standard errors. While these standard errors are correct, they reflect only within-study precision and do not capture uncertainty arising from between-study variability or reproducibility. As a result, these estimates can appear disproportionately precise, and are more vulnerable to study-specific bias and imprecision than pooled meta-analytic findings. These results should therefore be interpreted very cautiously and not weighed in the same way as the better-supported domains, . Similarly, family relationships and support and opportunities for education, employment and training were each based on only one study, and inferential statistics were not estimable.

The omnibus Q-test for between-group heterogeneity was not estimable as the variance-covariance matrix of the contrasts was not positive definite, which usually happens in the presence of several outcome domains with only one study and, in some cases, zero robust standard errors.

Country

To investigate potential differences in the effectiveness of CBT Skills programmes, subgroup analyses were conducted by country (Table 11).

Table 11: *Subgroup analysis on country*

Country	k	n	SMD (robust SE)	p-value	YEF impact rating
USA	108	21	−0.27 (0.08)	.004 **	High
Turkey	44	4	−1.38 (0.31)	.096 †	High
Nigeria	31	3	−1.22 (0.63)	.298	High
Spain	14	1	−2.63 (0.14)	n/a †	
Denmark	12	1	−0.23 (0.00)	n/a †	
Singapore	12	1	−1.29 (0.01)	n/a †	
England	9	2	−0.33 (0.11)	.195	High
Pakistan	7	1	−1.22 (0.06)	n/a †	
Netherlands	6	1	+0.04 (0.00)	n/a †	
Iceland	4	1	−0.23 (0.00)	n/a †	
Norway	4	1	−0.16 (0.02)	n/a †	
Between-group heterogeneity (Q_n test); 11 countries; Q_n = 1,041,498; p <.001 ***					

CBT Skills interventions showed mostly beneficial effects across countries, with negative subgroup estimates in nearly all cases, indicating better outcomes for participants receiving CBT Skills interventions than for controls. The clearest and most robust evidence came from the USA (k = 108 effect sizes, 21 studies), where CBT Skills interventions showed a statistically significant beneficial effect (SMD = −0.27, p = .004). This was by far the largest evidence base and therefore provides the most reliable country-specific estimate in this analysis.

Larger beneficial point estimates were also observed in Turkey (SMD = −1.38, p = .096), Nigeria (SMD = −1.22, p = .298), Spain (SMD = −2.63, p = .034), Singapore (SMD = −1.29, p = .007), and Pakistan (SMD = −1.22, p = .030). However, most of these country-specific estimates were based on only one study, and several had extremely small or near-zero standard errors. These results should therefore be interpreted very cautiously, as they may give a misleading impression of precision. England showed a smaller non-significant beneficial effect (SMD = −0.33, p = .195), while the Netherlands was the only subgroup with a positive estimate (SMD = +0.04, p < .001), suggesting no evidence of benefit in that country subgroup, though this too was based on only one study.

Finally, a Q-test for between-group heterogeneity indicated statistically significant differences across countries ($\chi^2(10)$ = 1,041,498, p < .001). However, this result should be treated with substantial caution. The extremely large test statistic is very likely being driven by sparse country categories, especially those based on a single study with near-zero estimated standard errors, rather than providing a stable indication of genuine between-country differences. In practical terms, although the omnibus test is formally significant, the only country subgroup with a clearly robust evidence base is the USA. The remaining country-specific findings are too sparse to support strong comparative conclusions about which countries show larger or smaller intervention effects.

Publication year

Subgroup meta-analyses were conducted by decade to explore variation in effect sizes over time, given the wide range of publication years across included studies (Table 12).

Table 12: *Subgroup analysis on publication year*

Decade	k	n	SMD (robust SE)	p-value	YEF impact rating
1980 - 1989	15	3	−0.48 (0.17)	.132	High
1990 - 1999	10	4	−0.98 (0.36)	.096 †	High
2000 - 2009	54	11	−0.05 (0.14)	.740	Small
2010 - 2019	106	10	−1.01 (0.25)	.008 **	High
2020 - 2026	66	9	−1.05 (0.48)	.070 †	High
Between-group heterogeneity (Q_n test); 5 decades; Q_n = 4.32; p<.002 **					

CBT Skills interventions showed beneficial effects in every decade subgroup, with all estimates in the negative direction, indicating better outcomes for participants receiving CBT Skills interventions than for controls. The clearest evidence was observed for studies conducted in the 2010s (k = 106 effect sizes, 10 studies), which showed a statistically significant beneficial effect (SMD = −1.01, p = .008). This subgroup also had the largest evidence base.

A similarly large beneficial effect was observed for studies from the 2020s (SMD = −1.05), although this was only marginally significant (p = .070). Studies from the 1990s also showed a relatively large beneficial effect (SMD = −0.98, p = .096), again with only marginal significance. In contrast, the 2000s subgroup showed little evidence of benefit overall (SMD = −0.05, p = .740), while studies from the 1980s showed a moderate but non-significant beneficial effect (SMD = −0.48, p = .132).

Finally, a Q-test for between-group heterogeneity indicated that the effects of CBT Skills interventions differed significantly across decades. A robust Wald chi-

square test showed significant between-group heterogeneity ($\chi^2(4) = 4.32$, p = .002), suggesting that the estimated effectiveness of CBT Skills interventions has varied over time. In practical terms, this pattern suggests stronger effects in the 1990s, 2010s, and 2020s, and much weaker effects in the 2000s, although differences in study design, implementation, and outcome measurement across decades may also contribute to this variation.

Age band

Subgroup meta-analysis was conducted to examine whether the effectiveness of CBT Skills interventions varied by the age range of participants. Studies were grouped into four categories based on the available data. These included Childhood and preadolescence (ages 5–11); Early adolescence (ages 12–15); Middle/late adolescence (ages 16+); and Mixed when studies had a broader age range of participants beyond these categories (Table 13).

Table 13: *Subgroup analysis on developmental age band*

Age range	k	n	SMD (robust SE)	p-value	YEF impact rating
Childhood and preadolescence (ages 5–11)	83	11	−0.82 (0.39)	.068 †	High
Early adolescence (ages 12–15)	62	8	−0.79 (0.53)	.232	High
Middle/late adolescence (ages 16+)	44	5	−0.79 (0.28)	.082 †	High
Mixed	62	13	−0.76 (0.31)	.047 *	High
Between-group heterogeneity (Q_n test); 4 categories; $Q_n = 0.005$; p = .999					

CBT Skills interventions showed beneficial effects across all age groups, with all subgroup estimates in the negative direction, indicating better outcomes for participants receiving CBT Skills interventions than for controls. The subgroup with

mixed-age samples (k = 62 effect sizes, 13 studies) showed a statistically significant beneficial effect (SMD = −0.76, p = .047). Childhood and preadolescence (ages 5–11) also showed a beneficial effect of similar magnitude (SMD = −0.82), although this was only marginally significant (p = .068). Likewise, middle/late adolescence (ages 16+) showed a beneficial effect (SMD = −0.79, p = .082) that was also marginally significant.

The estimate for early adolescence (ages 12–15) was very similar in magnitude (SMD = −0.79), but this subgroup did not reach statistical significance (p = .232), reflecting greater uncertainty rather than a meaningfully different effect size. Overall, the subgroup-specific estimates were remarkably consistent, all clustering around an SMD of about −0.76 to −0.82.

Finally, a Q-test for between-group heterogeneity indicated that the effects of CBT Skills interventions did not differ significantly across age groups. A robust Wald chi-square test showed no evidence of between-group heterogeneity ($\chi^2(3) = 0.005$, p = .999), suggesting that intervention effects were highly similar regardless of participant age range. In practical terms, this indicates that CBT Skills interventions appear to be similarly effective across childhood, adolescence, and mixed-age samples.

Ethnicity

To explore whether programme effectiveness varied by ethnicity composition, studies were grouped into five categories based on reported sample characteristics: 100% Black and Global Majority, majority white (>85% white participants), some diversity (15–49% Black and Global Majority participants), predominantly Black and Global Majority (≥50% Black and Global Majority participants), and ethnicity not stated (Table 14).

Table 14: *Subgroup analysis on ethnicity*

Ethnicity composition	K	n	SMD (robust SE)	p-value	YEF impact rating
Not stated	114	16	−1.15 (0.28)	.005 **	High

Diverse / balanced	57	11	−0.31 (0.05)	<.001 ***	High
100% of the sample is Black and Global Majority	37	4	−1.23 (0.49)	.144	High
Majority white sample	28	4	+0.14 (0.13)	.374	Harmful
Some diversity	15	3	−0.55 (0.20)	.134	High
Between-group heterogeneity (Q_n test); 5 categories; Q_n = 6.30; p<.001 ***					

CBT Skills interventions showed varying effects across ethnicity composition subgroups. The clearest evidence was observed in studies where ethnicity was not stated (k = 114 effect sizes, 16 studies), which showed a statistically significant beneficial effect (SMD = −1.15, p = .005). A statistically significant beneficial effect was also observed for diverse / balanced samples (k = 57, 11 studies), although the magnitude was smaller (SMD = −0.31, p < .001).

A larger beneficial point estimate was observed for studies in which 100% of the sample was Black and Global Majority (SMD = −1.23), but this effect was not statistically significant (p = .144), likely reflecting the smaller evidence base and greater imprecision. Some diversity also showed a beneficial but non-significant effect (SMD = −0.55, p = .134). In contrast, the majority white sample subgroup showed no evidence of benefit, with a small and non-significant estimate (SMD = +0.14, p = .374)..

Finally, a Q-test for between-group heterogeneity indicated that the effects of CBT Skills interventions differed significantly across ethnicity composition categories. A robust Wald chi-square test showed significant between-group heterogeneity ($\chi^2(4) = 6.30$, p < .001), suggesting that ethnicity composition may help explain variation in intervention effects. In practical terms, this indicates that the effectiveness of CBT Skills interventions may not be uniform across samples

with different ethnic compositions, although the largest body of evidence remains concentrated in studies where ethnicity was not stated.

Gender

To investigate whether the effectiveness of CBT Skills programmes varied according to the gender composition of the study samples, subgroup analyses were conducted across three categories: studies with only male participants, studies with majority male samples, and mixed samples (Table 15).

Table 15: *Subgroup analysis on gender*

Gender composition	k	n	SMD (robust SE)	p-value	YEF impact rating
Mixed sex	146	18	−1.00 (0.27)	.005 **	High
Majority male	70	12	−0.60 (0.17)	.008 **	High
All male	35	7	−0.27 (0.19)	.223	High
Between-group heterogeneity (Q_n test); 3 categories; Q_n = 2.49; p=.083 †					

CBT Skills interventions showed beneficial effects across all gender composition subgroups, with all subgroup estimates in the negative direction, indicating better outcomes for participants receiving CBT Skills interventions than for controls. The clearest evidence was observed in mixed-sex samples (k = 146 effect sizes, 18 studies), which showed a statistically significant beneficial effect (SMD = −1.00, p = .005). A statistically significant beneficial effect was also observed in majority male samples (k = 70, 12 studies), although the magnitude was somewhat smaller (SMD = −0.60, p = .008).

In contrast, all-male samples also showed a beneficial point estimate (SMD = −0.27), but this effect was not statistically significant (p = .223), suggesting weaker or more uncertain evidence of benefit in that subgroup.

Finally, a Q-test for between-group heterogeneity provided only marginal evidence that the effects of CBT Skills interventions differed across gender composition categories. A robust Wald chi-square test indicated borderline between-group heterogeneity ($\chi^2(2) = 2.49$, $p = .083$). In practical terms, this suggests that CBT Skills interventions may be somewhat more effective in mixed-sex and majority male samples than in all-male samples, but the evidence for differences between groups is not strong enough to draw firm conclusions.

Multiple meta-regressions (models 1-5)

Following these ten subgroup analyses, we ran five separate random-effects meta-regression models with robust variance estimation to examine pre-specified moderators: (1) quality, (2) setting, (3) intervention features, (4) population characteristics, and (5) outcome characteristics. See Appendix 1 for an overview of the methods used in this section and Appendix 3 for a list of the studies that provided data for these analyses.

Moderator Analysis 1 – Quality Moderators

The characteristics in model 1 include:

- Study design
- Unit of allocation into group[2]
- Method of allocation into group
- Quality appraisal as assessed by the YEF-EQA tool

Table 16: *Meta-regression (with robust inference) on quality moderators*

Moderator	β (estimate)	SE	t	df	p	95% CI
Intercept	−0.5270	0.426	−1.24	6.89	0.256	−1.54, 0.48
Design = RCT (vs reference)	−1.2055	4.139	−0.29	1.58	0.805	−24.42, 22.01

[2] This was dropped from the meta-regression model as only one level, i.e., all studies allocated at the individual level

Method of allocation = random (vs reference)	1.0273	4.132	0.25	1.11	0.842	−40.73, 42.78
Study quality (YEF EQA) = Low	0.7742	0.531	1.46	5.92	0.196	−0.53, 2.08
Study quality (YEF EQA) = Moderate	−0.0792	0.306	−0.26	4.23	0.808	−0.91, 0.75

Note. Values are from a multivariate random-effects meta-regression estimated using REML (k = 251), clustering on study ID to account for non-independence of multiple effect sizes within studies. Statistical inference is based on CR2-robust standard errors with Satterthwaite small-sample t-tests.

The estimated between-study variance component was $\tau^2 = 1.05$ (SD = 1.02), indicating substantial between-study heterogeneity. Residual heterogeneity remained highly significant, QE(df = 246) = 6492.87, p < .0001, showing that the moderators included in the model left substantial variance in effect sizes unexplained. . The omnibus test of moderators from the REML model was not statistically significant, QM(df = 4) = 5.89, p = .208, and none of the individual moderators were statistically significant in the robust analyses. In practical terms, this suggests that study design, method of allocation, and study quality did not meaningfully explain heterogeneity in effect sizes in this model.

The robust degrees of freedom for design and method of allocation were very small, and their confidence intervals were extremely wide, indicating substantial imprecision in those estimates.

Moderator Analysis 2 – Setting-level Moderators

The characteristics in model 2 include:

- Country of intervention
- Setting of intervention

Table 17: *Meta-regression (with robust inference) on setting-level moderators*

Moderator	β (estimate)	SE	t	df	p	95% CI

Intercept	−0.2180	1.031	−0.21	1.00	0.867	−13.32, 12.88
Country = England (vs reference)	0.3881	1.082	0.36	2.95	0.744	−3.09, 3.87
Country = Iceland (vs reference)	0.5261	1.107	0.48	2.45	0.674	−3.49, 4.54
Country = Netherlands (vs reference)	0.7886	1.087	0.73	2.12	0.540	−3.65, 5.22
Country = Nigeria (vs reference)	−1.2934	1.515	−0.85	3.02	0.456	−6.10, 3.51
Country = Norway (vs reference)	0.5518	1.411	0.39	3.02	0.722	−3.92, 5.03
Country = Pakistan (vs reference)	−0.2389	9.302	−0.03	3.01	0.981	−29.76, 29.29
Country = Singapore (vs reference)	−0.4208	11.844	−0.04	3.00	0.974	−38.11, 37.27
Country = Turkey (vs reference)	0.2883	1.886	0.15	2.73	0.889	−6.06, 6.64
Country = USA (vs reference)	0.2273	1.057	0.22	1.83	0.851	−4.75, 5.21
Country = Spain (vs reference)	−2.0789	23.305	−0.09	2.96	0.935	−76.89, 72.73
Intervention setting = Computerised / digital (vs reference)	0.3709	0.352	1.05	1.56	0.428	−1.64, 2.38

Intervention setting = Justice / custodial / community supervision setting (vs reference)	−0.3698	0.379	−0.98	2.23	0.423	−1.85, 1.11
Intervention setting = Residential care / social care setting (vs reference)	−0.7501	0.327	−2.30	2.31	0.132	−1.99, 0.49
Intervention setting = School / education setting (vs reference)	−0.5328	0.310	−1.72	1.32	0.287	−2.79, 1.72

Note. Denmark was the reference category for country, and community-based clinical/mental health service was the reference category for intervention setting. All coefficients for country and setting represent differences from these reference categories

The estimated between-study variance component was $\tau^2 = 1.69$ (SD = 1.30), indicating substantial variation across studies. A test for residual heterogeneity remained highly significant ($QE(df = 236) = 2665.44$, $p < .0001$), showing that the setting-level moderators included did not fully account for heterogeneity across effects. The omnibus test of moderators was not statistically significant ($QM(df = 14) = 12.70$, $p = .550$), indicating that, taken together, country and intervention setting did not significantly explain variation in effect sizes.

In simple terms, the non-significance of the moderators indicates that, in this model, neither the country in which the intervention took place nor the intervention setting meaningfully explained heterogeneity in effect sizes. Although the coefficient for residential care / social care settings was somewhat larger in magnitude than the other setting contrasts , it was not statistically significant, and its confidence interval included zero.

Moderator Analysis 3 – Intervention-level Moderators

The characteristics in model 3 include:

- Intervention intensity
- Intervention duration
- Intervention provider

- Training for intervention providers
- Delivery method
- Family sessions or components in addition to the CBT protocol

Table 18: *Meta-regression (with robust inference) on intervention-level moderators*

Moderator	β (estimate)	SE	t	df	p	95% CI
Intercept	1.1462	1.204	0.95	3.37	0.404	−2.46, 4.75
Duration = 1 week (and 1 day) to 1 month	−0.6526	0.850	−0.77	10.40	0.460	−2.54, 1.23
Duration = 3 months (and 1 day) to 6 months	0.1724	0.931	0.19	4.44	0.861	−2.32, 2.66
Duration = 6 months (and 1 day) to 1 year	0.3309	3.951	0.08	4.38	0.937	−10.28, 10.94
Intensity = Low intensity / brief interventions	0.2763	1.624	0.17	5.27	0.871	−3.84, 4.39
Intensity = Moderate intensity / standard interventions	−0.1016	1.100	−0.09	3.22	0.932	−3.47, 3.27
Implementer training = Training inferred as implementer developed the programme	−0.1608	0.670	−0.24	6.34	0.818	−1.78, 1.46
Implementer training = Yes, training outlined	−0.7688	0.821	−0.94	7.42	0.379	−2.69, 1.15
Delivery method = Hybrid interventions (face-to-face + technology/media)	−0.4151	1.427	−0.29	4.91	0.783	−4.10, 3.27
Delivery method = Individual, face to face	0.8153	0.505	1.62	5.50	0.162	−0.45, 2.08
Delivery method = Individuals and small groups (5 per group), face to face	1.7166	3.346	0.51	5.09	0.629	−6.84, 10.27
Delivery method = Large groups (15+), face to face	0.4120	0.750	0.55	4.56	0.609	−1.57, 2.40

Delivery method = Medium groups (6–14), face to face	1.0784	0.744	1.45	6.03	0.197	−0.74, 2.90
Delivery method = Small groups (2–5), face to face	0.2197	1.239	0.18	7.08	0.864	−2.70, 3.14
Provider = Mental health / therapeutic professional	−1.4936	1.666	−0.90	6.08	0.404	−5.56, 2.57
Provider = Not stated	−1.9554	2.484	−0.79	3.19	0.486	−9.60, 5.69
Provider = Research / study staff	−1.7511	1.381	−1.27	6.18	0.251	−5.11, 1.61
Family involvement = Integrated joint family sessions (high involvement)	0.8519	1.601	0.53	3.77	0.625	−3.70, 5.41
Family involvement = No family component	−0.4582	0.784	−0.58	6.37	0.579	−2.35, 1.43
Family involvement = Parallel parent management training (moderate involvement)	0.4364	1.740	0.25	5.05	0.812	−4.02, 4.89

The estimated between-study variance component was $\tau^2 = 1.84$ (SD = 1.36), indicating substantial variation across studies. A test for residual heterogeneity remained highly significant ($QE(df = 231) = 2638.18$, $p < .0001$), showing that the intervention-level moderators included did not fully account for heterogeneity across effects. The omnibus test of moderators was not statistically significant ($QM(df = 19) = 8.94$, $p = .975$), indicating that, taken together, these intervention-level characteristics did not significantly explain variation in effect sizes.

In simple terms, the non-significance of the moderators indicates that, in this model, intervention duration, intensity, implementer training, delivery method, provider, and family involvement did not meaningfully explain heterogeneity in effect sizes. Although some coefficients were larger in magnitude than others, none were statistically significant in the robust analyses, and all confidence intervals included zero. Overall, this suggests that these intervention-level characteristics were not reliable predictors of effect-size differences across studies in this model.

Moderator Analysis 4 – Population-level Moderators

The characteristics in model 4 include:

- Age band
- Ethnicity
- Gender
- Socio-economic status

Table 19: *Meta-regression (with robust inference) on population-level moderators*

Moderator	β (estimate)	SE	T	df	p	95% CI
Intercept	−0.8505	0.821	−1.04	6.93	0.335	−2.80, 1.10
Age range = Early Adolescence (Ages 12–15)	0.9314	0.769	1.21	13.17	0.247	−0.73, 2.59
Age range = Middle/late Adolescence (Ages 16+)	0.0071	0.550	0.01	8.66	0.990	−1.24, 1.26
Age range = Mixed	−0.2311	0.343	−0.67	14.52	0.511	−0.97, 0.50
Ethnicity = 100% of the sample is Black and Global Majority	−1.4230	0.497	−2.86	5.85	0.030	−2.65, −0.20
Ethnicity = Diverse / Balanced	−0.3546	0.507	−0.70	5.88	0.511	−1.60, 0.89
Ethnicity = Not stated	−1.0238	0.584	−1.75	5.91	0.132	−2.46, 0.41
Ethnicity = Some Diversity	−1.3246	0.787	−1.68	6.69	0.138	−3.20, 0.55

Gender = Majority male	0.3231	0.490	0.66	9.11	0.526	−0.78, 1.43
Gender = Mixed sex	0.1032	0.512	0.20	9.75	0.844	−1.04, 1.25
SES = Low SES	0.5839	0.641	0.91	6.93	0.393	−0.94, 2.10
SES = Mixed SES	0.7464	0.912	0.82	5.37	0.448	−1.55, 3.04
SES = Not stated	1.0852	0.712	1.52	6.78	0.173	−0.61, 2.78

The estimated between-study variance component was τ^2 = 1.16 (SD = 1.08), indicating substantial variation across studies. A test for residual heterogeneity remained highly significant (QE(df = 238) = 2836.36, p < .0001), showing that the participant-level moderators included did not fully account for heterogeneity across effects. The omnibus test of moderators from the REML model was statistically significant (QM(df = 12) = 26.95, p = .008), suggesting that these participant-level characteristics were associated with variation in effect sizes overall.

At the level of individual moderators, however, robust analyses indicated that only one coefficient was statistically significant: studies in which 100% of the sample was Black and Global Majority had significantly lower effect sizes relative to the reference ethnicity category, β = −1.4230, SE = 0.497, t = −2.86, df = 5.85, p = .030, 95% CI (−2.65, −0.20).

Moderator Analysis 5. Outcome Moderators

The characteristics in model 5 include:

- YEF's outcome category
- Methods used to collect the outcome data

Table 20: **Meta-regression (with robust inference) on outcome-level moderators**

Moderator	β (estimate)	SE	t	df	p	95% CI
Intercept	−0.4619	0.299	−1.54	7.86	0.162	−1.15, 0.23
Outcome = Behavioural difficulties	0.1266	0.187	0.68	2.82	0.550	−0.49, 0.74
Outcome = Building and maintaining relationships	0.0736	0.187	0.39	2.77	0.722	−0.55, 0.70
Outcome = Bullying	0.0572	0.207	0.28	1.97	0.809	−0.85, 0.96
Outcome = Crime and offending	0.0595	0.317	0.19	1.52	0.873	−1.81, 1.93
Outcome = Criminal peers	0.1693	0.285	0.60	2.44	0.603	−0.87, 1.20
Outcome = Drug and alcohol use	0.1074	0.096	1.13	1.06	0.454	−0.96, 1.17
Outcome = Family relationships and support	−0.2444	0.224	−1.09	2.99	0.360	−0.96, 0.47
Outcome = General mental health	0.5530	0.316	1.75	2.52	0.196	−0.57, 1.68

Outcome = Helping others	1.1363	1.160	0.98	1.94	0.433	–4.00, 6.27
Outcome = Meaningful relationships	0.1709	0.256	0.67	1.15	0.613	–2.23, 2.57
Outcome = Opportunities for education, employment and training	0.3085	0.224	1.38	2.98	0.262	–0.41, 1.02
Outcome = Parenting practices	0.0541	0.212	0.25	4.03	0.812	–0.53, 0.64
Outcome = Regulating and managing emotions	−0.2790	0.196	−1.42	1.20	0.360	−1.97, 1.41
Outcome = School engagement	−0.1402	0.218	−0.64	4.03	0.556	–0.75, 0.46
Outcome = Self-esteem	0.2373	0.285	0.83	2.83	0.469	–0.70, 1.18
Outcome = Victim of crime	0.3296	0.752	0.44	1.20	0.727	–6.19, 6.85
Data source = Child self-report	−0.2398	0.149	−1.61	1.29	0.311	−1.38, 0.90
Data source = Clinician assessment/observation	−0.1825	0.273	−0.67	2.80	0.555	–1.09, 0.73

Data source = Parent reported	−0.2647	0.130	−2.04	1.19	0.257	−1.40, 0.87
Data source = Peer reported	−0.3000	0.118	−2.55	1.95	0.129	−0.82, 0.22
Data source = Teacher reported	−0.1892	0.143	−1.33	1.40	0.360	−1.14, 0.76

A multivariate random-effects meta-regression was estimated using REML, with robust standard errors clustered by study ID to account for dependence among multiple effect sizes from the same study. The estimated between-study variance was substantial (τ^2 = 1.15, SD = 1.07). Residual heterogeneity remained highly significant, QE(df = 229) = 3799.37, p < .0001, indicating that the included moderators did not account for most of the variability in effect sizes. The omnibus test of moderators was statistically significant, QM(df = 21) = 731.49, p < .0001, suggesting that outcome domain and data source were associated with variation in effect sizes overall. However, none of the individual coefficients were statistically significant in the robust analyses, and all robust confidence intervals included zero. Overall, these findings suggest that, once clustering and small-sample corrections were taken into account, there was no clear evidence that specific outcome domains or data-source categories meaningfully explained heterogeneity in effect sizes.

Publication bias

Publication-bias analyses were conducted on 37 study-level aggregated effect sizes. Egger's regression test indicated significant funnel plot asymmetry (z = -1.99, p = .046), suggesting possible small-study effects. In contrast, Begg's rank-correlation test was non-significant (Kendall's τ = -0.10, p = .398), providing weaker evidence of asymmetry.

The test of the selection model parameters was statistically significant (LRT(3) = 18.83, p < .001), suggesting that the probability of studies being observed may

have varied across p-value intervals and that selective reporting cannot be ruled out. However, the model was unstable, and the individual selection weights could not be estimated precisely. As illustrated in Figure 3, the estimated pattern of selection should therefore be interpreted cautiously.

Trim-and-fill analysis (Figure 4) estimated that 11 studies were missing on the right side of the funnel plot. After imputing these studies, the adjusted pooled effect was positive and no longer statistically significant ($\beta = 0.34$, SE = 0.20, $z = 1.74$, $p = .082$, 95% CI [-0.04, 0.73]). This suggests that the overall effect may be sensitive to possible publication bias and/or small-study effects.

A step-function[3] selection model was fitted to examine possible publication bias and/or selective reporting, using two-sided p-value intervals of ≤ .025, .025–.05, .05–.10, and > .10. The model was estimated using a random-effects model with ML estimation[4] based on 37 studies. The adjusted pooled effect remained negative and statistically significant ($\beta = -0.21$), indicating that the intervention effect remained in the favourable direction after allowing for possible selection.

Residual heterogeneity remained substantial in the selection model ($\tau^2 = 2.66$, $\tau = 1.63$), and the heterogeneity test remained highly significant (LRT(1) = 1575.34, $p < .0001$), indicating considerable between-study variation even after accounting for possible selection.

In simple terms, these findings suggest that the available published studies may not provide a fully balanced picture of the evidence. This is consistent with a wider pattern described in the literature whereby positive or statistically significant findings are more likely to appear in the published record, while harms and other unfavourable outcomes may be under-reported, which can make the available evidence look more favourable than it is in reality (Rosenthal, 1979).

[3] A step-function selection model is a publication-bias sensitivity analysis that divides studies into bands according to their p-values and estimates whether studies in some bands are more likely to be observed than others.

[4] ML refers to maximum likelihood estimation, a method used to estimate the model parameters by identifying the values that make the observed data most probable under the assumed model.

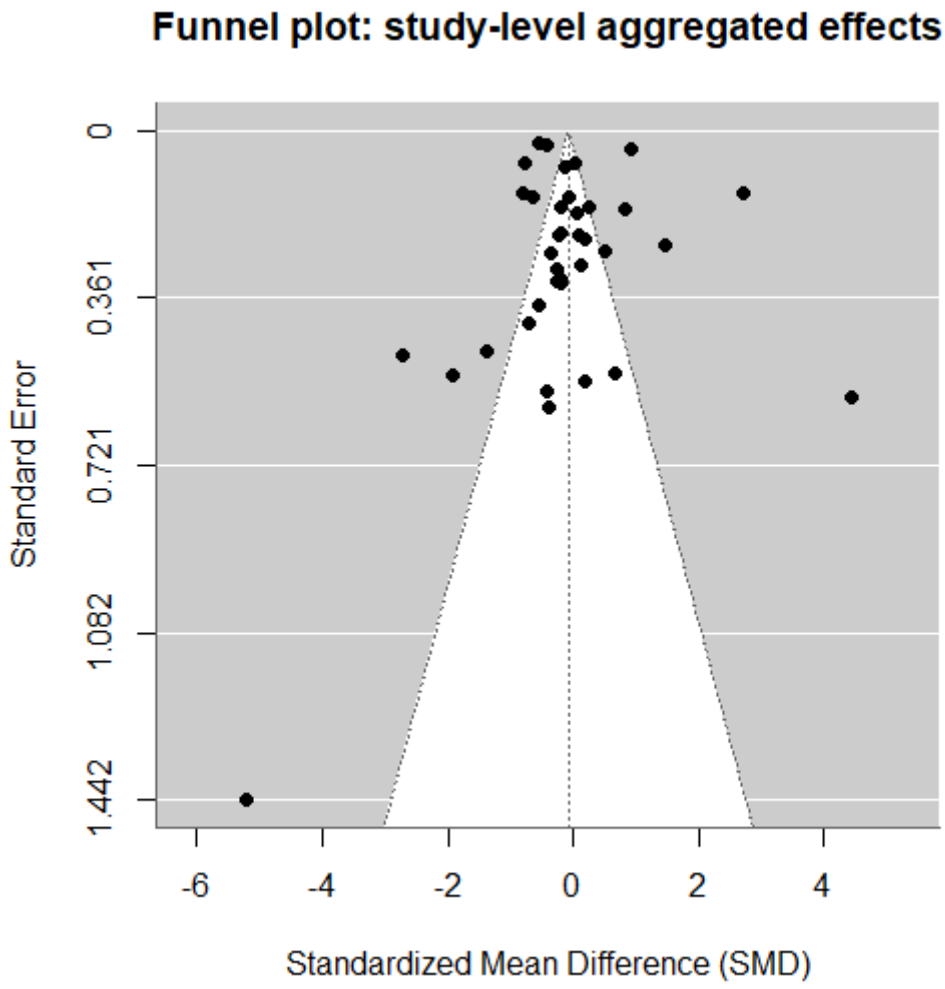

***Figure 3*: Funnel plot of individual studies**

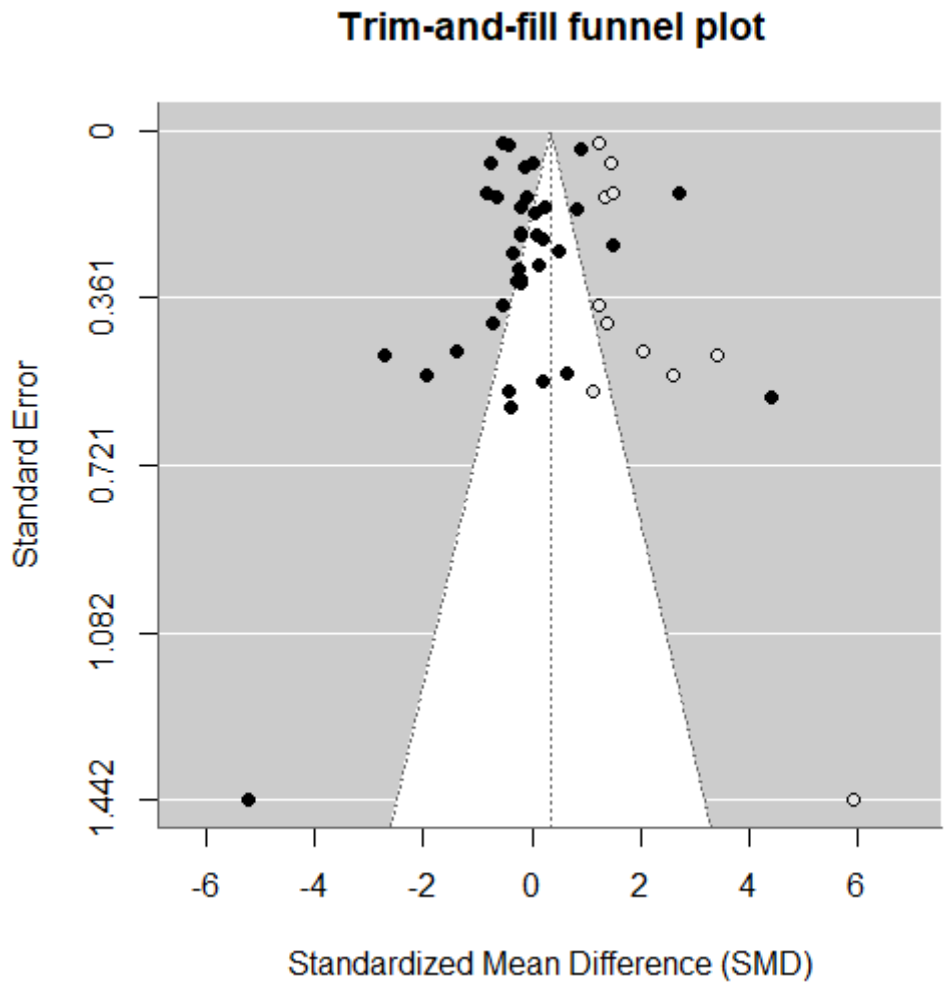

***Figure 4*: Trim and Fill Funnel plot of individual studies**

Sensitivity analyses

An examination of the studentized residuals revealed that several outcome-level residuals exceeded ±3 and may represent potential outlying observations in the context of this model. Specifically, one effect size from Karatas (2009) (z = 5.30), three effect sizes from Orgiles (2023) (z = -4.40, -3.67, and -3.36), and one effect size from Ime (2025) (z = -3.01) exceeded this threshold. Study-level influence diagnostics similarly suggested that a small number of studies were comparatively influential (Figure 5). Karatas (2009) was the most influential study,

with the largest Cook's distance (0.27) and the largest absolute DFBETA[5] for the intercept (0.69). Ime (2025) also appeared relatively influential, with the second-largest Cook's distance (0.23) and a comparatively large DFBETA for the intercept (-0.57).

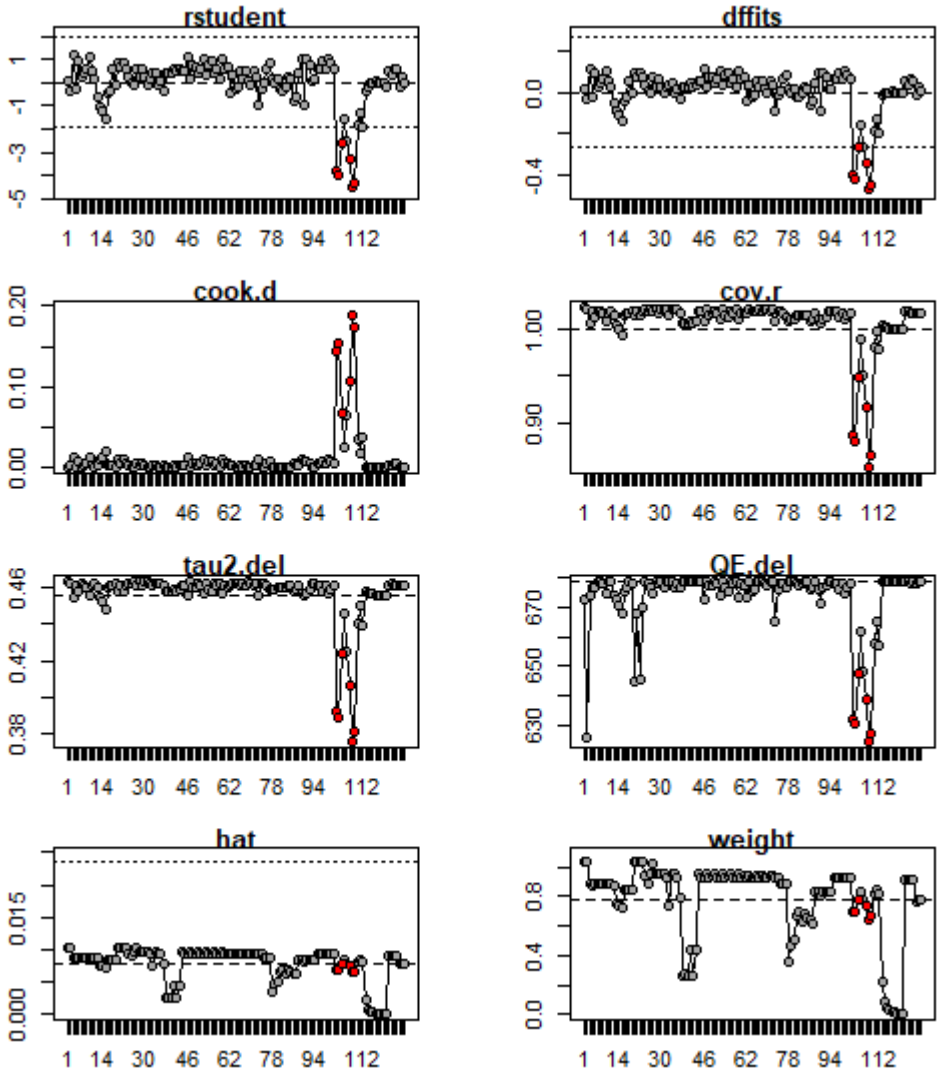

***Figure 5*: Influence diagnostics for individual studies**

Given this pattern, a leave-one-study-out analysis was conducted to evaluate the influence of individual studies on the pooled effect (Figure 6). Removal of

[5] In metafor, DFBETAS stands for "difference in betas" with beta meaning a model coefficient value. DFBETAS tells the reader how many standard deviations the estimated coefficient changes after excluding that study from the model fit.

each study in turn produced pooled effect estimates ranging from approximately -0.50 to -0.67. No single study reversed the direction of the effect, and all leave-one-study-out estimates remained statistically significant, indicating that the beneficial findings and conclusions of this meta-analysis were not unduly driven by any individual study. Omitting Karatas (2009) yielded the largest change in the pooled effect, increasing the magnitude of the effect to SMD = -0.67 (95% CI [-0.93, -0.41], $p < .001$) and substantially reducing residual heterogeneity ($\tau^2 = 0.58$). Omitting Ime (2025) produced the smallest-magnitude pooled effect, SMD = -0.50 (95% CI [-0.79, -0.20], $p < .001$), and also reduced residual heterogeneity ($\tau^2 = 0.75$). These findings suggest that both studies contributed disproportionately to the magnitude of the pooled effect and to the remaining heterogeneity, although neither was solely responsible for the overall findings.

To assess the robustness of the overall effect (SMD = -0.58) to quality decisions, a series of sensitivity analyses were conducted. First, analyses were repeated after excluding studies rated as 'low' quality. The overall pooled effect remained statistically significant and was somewhat larger in magnitude than in the main analysis (SMD = -0.77, SE = 0.16, $p < .001$), indicating that the primary findings were not driven by lower-quality studies. Second, the analysis was restricted to RCTs only, excluding quasi-experimental designs. The pooled effect estimate in this subset remained statistically significant and was comparable in magnitude to the main result (SMD = -0.61, SE = 0.20, $p = .004$), suggesting that study design did not materially influence the overall conclusion.

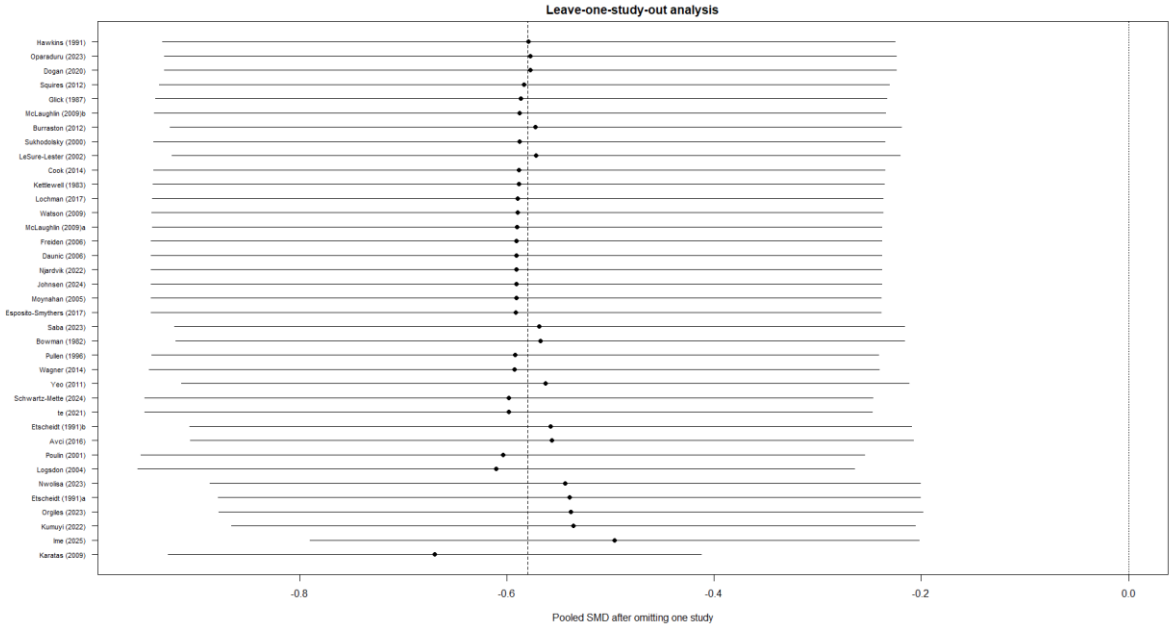

***Figure 6*: Leave-one-out sensitivity analysis**

Overall, the main conclusion appears robust: the pooled effect remained beneficial across all sensitivity analyses. However, the magnitude of the effect and the amount of unexplained heterogeneity were influenced most notably by Karatas (2009) and Ime (2025), and several individual effect sizes from Orgiles (2023) also appeared potentially outlying in the context of the model.

How Secure is the Evidence?

Violence outcomes

Our confidence in the findings on the prevention and reduction of violence is **Moderate.** The evidence for violence achieved an initial security rating of **Level 3.** This is based on 56 outcomes from 22 impact evaluations assessing the impact of CBT Skills programmes on children and young people

The studies included:

- 20 RCTs (3 high quality Type A studies; 15 moderate quality Type C studies and two low quality Type D studies)
- Two QEDs (a Moderate quality Type C study and a low quality Type D study)

Heterogeneity was high (I^2 = 95.6%) and moderator analyses were employed to investigate differences.

Crime and offending outcomes

Our confidence in the findings on crime and offending is **Very Low.** The evidence for crime and offending achieved an initial security rating of Level 1. The meta-analysis included 11 crime and offending related outcomes drawn from 4 studies assessing the impact of CBT Skills programmes on children and young people

The studies included:

- Two RCT's, of these, both rated as low quality (Type D)
- Two QED's, both rated as moderate quality (Type C)

Heterogeneity was high (I^2 = 68.8%), and moderator analyses were employed to investigate differences.

Overall CBT Skills programmes

Our confidence in the overall findings for CBT Skills is **Moderate.** The full effectiveness analysis achieved an overall security rating of **Level 3.** The combined evidence base consists of 37 impact evaluations, including 33 RCTs and 4 QEDs. Of these studies, four were rated as high quality (Type A), 24 as moderate quality (Type C), and nine as low quality (Type D). Heterogeneity was high (I^2 = 96.3%) in this model.

The subgroup analyses suggested that some characteristics may be associated with differences in effect size, with statistically significant between-group heterogeneity observed for participant inclusion criteria, intervention setting, family involvement, publication decade, and ethnicity composition, and borderline evidence for gender composition. By contrast, there was no clear evidence that effects differed reliably by programme intensity, delivery format, or age band. The multiple meta-regression analyses provided a more conservative test of these patterns. Of the five meta-regression models, the omnibus test was statistically significant for the population-level moderators model and for the outcome/domain and data-source model, suggesting that these sets of variables may explain some variation in effect sizes overall.

The results were nevertheless generally robust across sensitivity analyses and after applying robust variance estimation to account for dependent effect sizes. Taken together, these findings support confidence that CBT Skills programmes are likely to lead to beneficial effects, while also indicating that the magnitude of that effect varies across studies and that substantial unexplained heterogeneity remains. **The Level 3 evidence security rating was therefore retained.**

Who does it work for?

This narrative summary of equity-related outcomes supplements the meta-analysis by providing additional insights into how CBT Skills programmes may affect different demographic groups.

Four studies provided detail in relation to personal characteristics that help to explain for whom CBT Skills programmes work (Nwolisa, 2023; Oparaduru, 2017; Wagner et al., 2014; Watson, 2009). The studies covered gender, ethnicity, socioeconomic status and age. Studies took place in Nigeria (Nwolisa, 2023; Oparaduru, 2017), the UK (Watson, 2009) and the USA (Wagner et al., 2014). Using the YEF-EQA tool three studies were rated as moderate quality (Nwolisa, 2023; Oparaduru, 2017; Wagner et al., 2014) and one as low quality (Watson, 2009).

Studies where personal characteristics of the sample were described (e.g., gender) but not specifically related to outcomes of interest or do not contribute to the understanding of who CBT Sills programmes work for, have not been included in this section. No studies explored education, neurodiversity, care-experience, intersectionality or place of residence.

Gender

Four studies explored whether gender moderated the effects of CBT Skills interventions across different populations and settings (Oparaduru, 2017; Nwolisa, 2023; Wagner et al., 2014; Watson, 2009).

Three studies found no evidence that gender significantly influenced outcomes following CBT Skills programmes (Oparaduru, 2017; Wagner et al., 2014; Watson, 2009). Oparaduru (2017) reported no significant gender differences in post-test scores for maladaptive behaviour, self-esteem, or peer influence following cognitive restructuring and self-control interventions for secondary school students. Similarly, Wagner et al. (2014) found that gender did not moderate treatment effects in the Guided Self-Change intervention, with comparable reductions in alcohol use, drug use, and aggressive behaviour observed for males and females. Watson (2009) also found that gender was not a significant predictor of total Strengths and Difficulties Questionnaire (SDQ) scores or any behavioural and emotional subscales at follow-up in a primary school CBT anger

management intervention. In contrast, one moderate quality study (Nwolisa, 2023) identified significant gender differences in some outcomes, with female participants showing greater improvements in social anxiety-related thoughts and verbal aggression, although no gender differences were observed for depression or self-esteem. Overall, CBT Skills interventions appear to produce broadly similar outcomes for males and females across behavioural and cognitive measures. One study suggests gender differences emerge in specific emotional and aggression-related outcomes, indicating that while effects are generally equitable, certain domains may warrant gender-responsive attention.

Ethnicity

One moderate quality study explored whether ethnicity moderated the effectiveness of a school-based CBT-informed intervention for substance use and aggressive behaviour (Wagner et al., 2014). The study evaluated the Guided Self-Change (GSC) programme compared with standard care among a predominantly Black and Global Majority sample of high school students. Analyses found no evidence that ethnicity moderated treatment effects, with similar reductions in alcohol use, drug use, and interpersonal aggression observed for both Hispanic and African-American students. These findings suggest that the intervention produced comparable outcomes across different ethnicities included in the study. However, the analysis included only two ethnicities and did not assess wider racial, cultural, or linguistic diversity within the sample. While outcomes appeared comparable for Hispanic and African-American students, further research is needed to evaluate school-based CBT Skills programmes across a broader range of racial and ethnic populations.

Socioeconomic status

One low quality study explored whether socioeconomic status moderated the effectiveness of a CBT-based anger management intervention for primary school children (Watson, 2009). Socioeconomic deprivation was measured using eligibility for free school meals and the Index of Multiple Deprivation (IMD). The findings indicated that socioeconomic status was generally not a significant moderator of intervention outcomes, with IMD not predicting total Strengths and Difficulties Questionnaire (SDQ) scores or emotional, conduct, hyperactivity,

prosocial, and peer subscales. However, eligibility for free school meals was found to significantly predict outcomes on the emotional and hyperactivity subscales, suggesting some variation in outcomes associated with socioeconomic disadvantage. The study also noted that schools in the intervention group were located in more deprived areas than those in the wait-list control group, which may have influenced the results. Overall, evidence on socioeconomic status is limited and mixed with most outcomes showing no differential effects by deprivation, though some indicators of disadvantage may relate to emotional and behavioural outcomes. This underscores the need for future studies to use more sensitive, individual-level measures of socioeconomic status to clarify how deprivation influences access to, and effectiveness of, CBT-based interventions.

Age

One low quality study explored whether age moderated the effectiveness of a CBT-based anger management intervention for primary school children (Watson, 2009). Age was examined as a potential predictor of outcomes measured using the Strengths and Difficulties Questionnaire (SDQ). The findings indicated that age was generally not a significant predictor of total SDQ scores or emotional, conduct, hyperactivity, prosocial, and peer subscales following the intervention. However, age showed a significant moderating effect on changes in total SDQ scores between time 2 and time 3 for the wait-list control group, with younger children showing greater improvements over time compared to older children. These findings highlight the importance of considering developmental stage in CBT-based anger management interventions, while also noting the limited strength of the evidence and the need for further research to clarify how age influences treatment responsiveness.

What factors affect implementation?

Thirteen studies provided evidence related to implementation (Augustyniak et al., 2009; Cole et al., 2013; Cook et al., 2014; Daunic et al., 2006; Esposito-Smythers et al., 2017; Lochman et al., 2017; Logsdon, 2003; Orgilés et al., 2023; Pullen, 1996; Schwartz-Mette et al., 2024; Te Brinke et al., 2021; Wagner et al., 2014; Watson, 2009). Using the YEF-EQA, two of these papers were classified as high quality (Daunic et al., 2006; Lochman et al., 2017), seven as moderate quality (Cole et al.,

2013; Cook et al., 2014; Esposito-Smythers et al., 2017; Orgilés et al., 2023; Schwartz-Mette et al., 2024; Te Brinke et al., 2021; Wagner et al., 2014), three as low quality (Augustyniak et al., 2009; Logsdon, 2003; Watson, 2009), and one as very low quality (Pullen, 1996). Individual study details are available in Appendix 4.

Factors that influenced the implementation of CBT Skills programmes are organised using Proctor et al.'s (2011) Implementation Outcomes Framework. Appendix 5 highlights the availability of evidence according to each of Proctor's implementation outcomes. Some features of Proctor's framework were not evident in studies included in this review (sustainability).

Acceptability

Acceptability explores aspects of the intervention or change that children and young people find agreeable or satisfactory. Three studies, one high quality (Lochman et al., 2017), one moderate quality (Daunic et al., 2006), and one low quality study (Watson, 2009) examined the acceptability of CBT Skills programmes. Acceptability was reflected through participant satisfaction ratings, teacher perceptions, and qualitative feedback from children, parents, and facilitators.

Lochman et al. (2017) reported particularly strong evidence of acceptability for the Coping Power–Internet Enhanced (CP-IE) programme, a hybrid indicated prevention intervention designed for preadolescent youth displaying aggressive or disruptive behaviours. CP-IE is an adaptation of the evidence-based Coping Power programme that integrates traditional face-to-face cognitive-behavioural group sessions with an interactive web-based component. Satisfaction ratings were high among both children and parents at the end of the intervention. On a 0–7 scale, children reported a mean overall satisfaction rating of 6.63, while parents reported a mean of 6.83, indicating that the programme was viewed very positively by both groups. The face-to-face components were especially well received, with children rating the helpfulness of the in-person child sessions at 6.40, while parents rated the parent meetings at 6.72. Ratings of the online component were also positive, although somewhat lower, particularly among parents. Children rated the website's helpfulness at 6.08, and ease of access at 5.70, whereas parents rated the website's helpfulness at 4.61, and ease of access

at 5.29. These findings suggest that the hybrid format was broadly acceptable, with especially strong endorsement of the in-person elements, and more mixed, although still positive, views of the web-based content among parents.

Daunic et al. (2006) examined the acceptability of Tools for Getting Along (TFGA), a classroom-based cognitive social problem-solving curriculum, primarily through teacher reports. Teachers gave positive ratings to the programme's appeal and utility for students, with a mean score of 4.18 out of 5, indicating that the programme was viewed as engaging and beneficial for children. Perceptions of its broader value and usefulness in addressing behaviour were also favourable, with an overall scale mean of 3.86 out of 5. Follow-up discussions supported these findings, as many teachers indicated that they would use the programme again, and recommend it to other school staff. Although this evidence comes from the perspective of those delivering the intervention rather than the children themselves, it still suggests that the curriculum was considered worthwhile and practical within the classroom setting.

Watson (2009) provides a more mixed picture of acceptability, drawing on qualitative interviews with children, parents, and facilitators following an anger management intervention for primary school children. Children generally described positive changes after the programme, often reporting that they were better able to control their anger and use calming strategies such as counting. Positive views of the group were common: nine children in the treatment group and all eleven children in the control group expressed favourable feelings about participation, while thirteen children specifically emphasised that the group was enjoyable and fun. Children and facilitators also responded positively to the use of rewards, and facilitators highlighted the importance of varied activities in maintaining children's enjoyment. However, not all stakeholders were as consistently positive. Some parents and facilitators reported occasions where behaviour appeared to worsen, suggesting that the intervention was not experienced in the same way across all groups. This makes the evidence for acceptability less clear than in the other studies, even though children's own accounts were largely favourable.

Appropriateness

Appropriateness refers to the perceived fit or relevance of an intervention to the given context or problem. It can include discussion of adaptations that are made to improve the intervention's fit within the context and the perceived usefulness of the intervention. Five studies, one high quality study (Lochman et al., 2017), two moderate quality studies (Daunic et al., 2006; Orgilés et al., 2023), one low quality study (Logsdon, 2003) and one very low quality study (Pullen, 1996) examined the appropriateness of CBT Skills programmes. Across these studies, appropriateness was reflected in two main ways: how well interventions matched the developmental needs of children and young people, and how well they fitted the practical demands of the settings in which they were delivered. While several studies described interventions that were deliberately adapted to improve fit, they differed in their methods, with some focusing on accessibility and delivery context, while others focused on developmental suitability and engagement.

A number of studies emphasised contextual fit, particularly in relation to school or service delivery. Orgilés et al. (2023) evaluated a self-applied computerised version of Super Skills for Life (SSL) for children with emotional problems, and framed the digital format as a way of improving access for children who may not be able to engage with face-to-face programmes. The intervention retained the core objectives of the original SSL but adapted activities into a computerised format, using two digital characters to guide the child throughout the programme without continuous therapist supervision. Therapist involvement was not removed entirely; clinicians had the option to monitor performance data and remained available for parents' questions. However, the extent to which clinicians monitored data, and how frequently parents made use of this support, was not reported. The programme also used shorter, more frequent sessions, which may have been better suited to children's routines and attention spans. In this case, appropriateness was tied closely to accessibility, with the computerised format intended to reduce barriers to treatment and potentially shorten waiting times.

Lochman et al. (2017) also focused on adapting delivery format, but in a way that was specifically responsive to the elementary school context. The Coping Power–Internet Enhanced (CP-IE) programme was designed to maintain the core content of the original intervention while shifting a substantial part of the

psychoeducational material online, resulting in a reduction of approximately 60% in face-to-face sessions. This made the programme easier to fit within school schedules and reduced burden on staff, families, and students. Sessions were scheduled flexibly in consultation with teachers, and schools provided computer access for online activities where needed. The platform also included animated videos, quizzes, worksheets, and rewards, with separate access for children, parents, teachers, and counsellors. Like Orgilés et al. (2023), this study used digital adaptation to improve fit, but here the emphasis was aligning the intervention with the organisational demands of a school setting, rather than on broadening access.

Daunic et al. (2006) similarly demonstrated strong fit with the school environment through a more traditional classroom-based format. Tools for Getting Along (TFGA) was designed to be embedded within routine teaching, with lessons lasting around 30 minutes delivered twice weekly, allowing teachers to incorporate the programme without major disruption to academic instruction. Teacher ratings on the appropriateness and ease-of-use subscale averaged 3.99 out of 5, suggesting that the curriculum was viewed as reasonably compatible with classroom routines and suitable for students' developmental level. The use of familiar teaching methods, such as group discussion, paired work, worksheets, and role-play, further supported its compatibility with everyday classroom practice. Taken together, Lochman et al. (2017) and Daunic et al. (2006) both suggest that school-based interventions were perceived as more suitable when their structure was deliberately built around existing routines and staff practices.

Other studies highlighted appropriateness through developmental adaptation relative to delivery context alone. Logsdon (2003) evaluated a school-based social-cognitive-behavioural group intervention for aggressive boys aged 8 to 11, and showed how pilot testing was used to refine the intervention for this age group. Early findings suggested that some original activities were too abstract or not well suited to younger children, so these were removed and replaced with more structured, concrete, and interactive materials, including art activities, card-based scenarios, and games. These changes reportedly improved attention, reduced conflict, and increased participation. The intervention also focused on practical social skills such as emotional identification, sharing, and conversational abilities, which were more clearly aligned with the children's developmental

capacities. Compared with the school-fit emphasis seen in Lochman et al. (2017) and Daunic et al. (2006), Logsdon (2003) gives more weight to tailoring the content itself to children's cognitive and social level.

Pullen (1996) offers the clearest example of where the appropriateness of the intervention could be improved. In evaluating the Reasoning and Rehabilitation programme within Juvenile Intensive Supervision Probation, professionals raised concerns about whether the intervention's structure matched the needs of the young people involved. The most consistent issue was programme length: some respondents indicated that the programme should be reduced by half, while others suggested shortening sessions from two hours to 60–90 minutes. Staff also reported that young people often lost interest halfway through the programme, indicating that the intensity and duration were not well aligned with participants' engagement capacity. In contrast to the other studies, which generally described deliberate efforts to adapt delivery or materials to increase fit, Pullen (1996) highlights what happens when those adjustments appear insufficient or absent.

Across the studies, there is a clear contrast between interventions that were actively adapted to suit their context or population, and those for which the fit appeared more problematic. Orgilés et al. (2023), Lochman et al. (2017) and Daunic et al. (2006) all emphasised practical compatibility, whether through digital delivery, reduced face-to-face demands, or brief classroom-friendly sessions. Logsdon (2003) focused more strongly on developmental fit, showing the value of adjusting activities to children's abilities and attention spans. However, Pullen (1996) suggests that when intervention length and structure are not well matched to participants, appropriateness may be undermined even if the programme itself is evidence-informed.

Adoption

Adoption concerns the decision or action to employ an intervention or implementation target. It also refers to the uptake of an intervention provided by services and communities. One low quality study explored the adoption of CBT Skills programmes (Logsdon, 2003). Logsdon (2003) evaluated a school-based social-cognitive-behavioural group intervention aimed at reducing aggressive

behaviour and preventing later delinquency among boys aged 8 to 11. Adoption was notably limited in this study and represents a key implementation challenge. Despite efforts to recruit a broad range of sites across multiple states including school districts in New York, Vermont, Oregon, and Washington, as well as a tribal council, only two school systems agreed to participate: one in the Seattle area and one in the Woodstock, New York area. This low level of uptake indicates substantial barriers at the organisational level. The authors note that there was resistance from school officials; however, the specific nature of these barriers was not described in detail. As such, while it is evident that challenges to adoption existed, the precise organisational, structural, or attitudinal factors underlying this resistance remain unclear. Institutional and administrative constraints likely contributed to the limited adoption. These findings underscore the need to secure organisational buy-in and address contextual barriers early, as willingness to adopt cannot be assumed even when interventions target salient issues like childhood aggression.

Feasibility

Feasibility refers to the extent to which an intervention can be successfully implemented within a specific setting, with emphasis on its practicality and the ability to deliver it effectively in the target environment. This includes evidence of practicality or utility, and other indicators demonstrating that the intervention can be carried out as intended in real-world conditions. Four studies, one high quality study (Lochman et al., 2017), two moderate quality studies (Augustyniak et al., 2009; Esposito-Smythers et al., 2017) and one low quality study (Watson, 2009) examined the feasibility of CBT Skills programmes. Feasibility was reflected in how well interventions could be fitted into schools or services, the extent to which delivery procedures were manageable, and whether practical barriers limited implementation.

Esposito-Smythers et al. (2017) present a more mixed picture of feasibility for the family-based alcohol, self-harm and HIV prevention (ASH-P) programme. The intervention itself was quite intensive. Families in the ASH-P condition attended a 12-hour workshop delivered over two consecutive weekends, which included an 8-hour session followed by a 4-hour session as well as a 2-hour booster session two weeks later. Although the study successfully recruited 81 eligible families, the

authors noted that this format may have reduced the practicality of the programme. Requiring adolescents and parents to commit to two weekends of workshop attendance created a substantial time burden, which the authors identified as a likely challenge for recruitment and retention. They suggested that a shorter format, such as a single-day workshop, integration into routine clinical care, or technology-assisted delivery, might be easier to implement in practice. This suggests that the intervention was deliverable within the study, but that its intensity and scheduling demands may limit feasibility in usual service contexts.

Lochman et al. (2017) reported stronger evidence for the feasibility of Coping Power–Internet Enhanced (CP-IE) in elementary school settings. A key feature supporting implementation was the reduced intervention burden compared with the standard Coping Power programme. CP-IE required 12 child sessions and 7 parent sessions, compared with 34 child sessions and 16 parent sessions in the original version. This shortened format made it easier to fit sessions into the school day and reduced demands on staff, families, and schools. Sessions were successfully scheduled during lunch periods, early mornings, and afternoon slots, showing that the programme could be accommodated within routine school structures. The online component also supported implementation by shifting some psychoeducation, skills practice, and behavioural monitoring outside face-to-face sessions. Teachers reportedly found the web-based goal-monitoring system convenient, and schools were able to provide computer access for children who needed it. Ongoing supervision for trained interventionists, along with attendance tracking and session documentation, further suggests that the programme could be delivered and monitored successfully in practice.

Augustyniak et al. (2009) also provide support for school-based feasibility. The Prepare Curriculum was delivered in 10 groups across 10 school districts using a 10-week protocol, which the authors described as being specifically selected because it was well suited to school settings. The group format was seen as an efficient way to work within limited school resources while reaching more students. This suggests the intervention was viewed as manageable within school constraints, as the psychoeducational group model enabled support to be delivered efficiently to multiple young people at once.

By contrast, (Watson, 2009) highlights several practical difficulties that affected feasibility in school settings. Facilitators reported challenges organising the intervention, including problems finding suitable space in schools, scheduling sessions, and arranging for a co-facilitator to attend. Behaviour within the group also created implementation challenges, as both facilitators and children described disruption during sessions and difficulties between some group members. These issues suggest that, although the intervention could be delivered, doing so consistently and easily within the school environment was not always straightforward.

Across the studies, a clear contrast emerges between interventions that were designed to fit existing service contexts and those that placed heavier demands on participants or staff. Lochman et al. (2017) and Augustyniak et al. (2009) suggest that school-based group interventions can be implemented relatively successfully when they are concise and aligned with available resources. However, Watson (2009) shows that school delivery can still be undermined by everyday logistical and behavioural difficulties. Esposito-Smythers et al. (2017) differs from the school-based studies in that the main challenge was not the institutional setting, but the burden placed on families by an intensive weekend workshop format.

Fidelity

Fidelity refers to the degree to which an intervention was delivered as intended. Eleven studies, two high quality studies (Daunic et al., 2006; Lochman et al., 2017), six moderate quality studies (Cole et al., 2013; Cook et al., 2014; Esposito-Smythers et al., 2017; Schwartz-Mette et al., 2024; Te Brinke et al., 2021; Wagner et al., 2014), two low quality studies (Augustyniak et al., 2009; Watson, 2009) and one very low quality study (Pullen, 1996) examined the fidelity of CBT Skills programmes. Across these studies, intervention fidelity was generally supported through manualised protocols, structured training, and ongoing supervision. However, the studies differed in the extent to which fidelity was directly assessed, with some providing detailed observational data, and others relying more on broader descriptions of implementation procedures.

One clear pattern across the studies is the widespread use of manualised intervention protocols to promote consistency. Manuals were utilised in the Prepare Curriculum, Becoming a Man (BAM), ASH-P, Coping Power – Internet Enhanced (CP-IE), the FRIENDS Resilience programme, and in Guided Self-Change, (Augustyniak et al., 2009; Cook et al., 2014; Esposito-Smythers et al., 2017; Lochman et al., 2017; Schwartz-Mette et al., 2024; Wagner et al., 2014). Across school-based, family-based, and group interventions, the use of manuals appears to have been a common strategy for standardising delivery. Yet the strength of fidelity evidence depended less on the presence of a manual alone and more on whether studies paired manuals with direct monitoring of adherence and competence.

The most convincing evidence came from studies that combined the use of manuals with session recording, structured coding, and clear quantitative indicators of adherence. Esposito-Smythers et al. (2017) provides particularly strong evidence in this regard. Seven master's-level interventionists completed a two-day training involving psychoeducation and supervised practice before delivering the ASH-P workshops. All workshops were audiotaped, and 70% of audio (95 of 140 hours) from 7 of 11 randomly selected workshops was reviewed. These reviews showed 100% adherence to core intervention components, while therapist competence was rated as "well" or "very well" in over 90% of observed delivery. This combination of structured training, extensive audio review, and assessment of both adherence and competence gives provides strong evidence that the intervention was implemented as intended.

A similarly rigorous approach was reported by Schwartz-Mette et al. (2024) in the evaluation of the FRIENDS Resilience programme. Delivery was supported by substantial training, with the lead author completing a 5-day, 20-hour training with programme developers to become a Licensed FRIENDS Trainer; facilitators then received approximately 16 hours of training. Fidelity procedures included scripted session plans, post-session checklists, weekly review meetings, and direct observation of 30% of sessions in each cohort. Reliability between raters was exceptionally high, with ICCs of 1.00 between external raters and 0.98–1.00 between facilitators and external raters. Observed implementation was also very strong, with 99.2% of session objectives met, 98.6% of scheduled activities completed, and 100% of at-home activities assigned. No significant cohort

differences were identified in protocol integrity, suggesting that delivery remained highly consistent across different facilitator groups.

Te Brinke et al. (2021) evaluated different ways of sequencing cognitive and behavioural modules ("Think Cool; Act Cool") to determine if the intervention could reduce externalising problems and increase emotion regulation. They reported detailed fidelity data, including extending the analysis beyond adherence alone to consider whether the two modules remained distinct in practice. All sessions were audiotaped, and 104 sessions (23.5%) were randomly selected for coding. General session content received ratings of 3 or 4 on a 4-point scale in 76.89% of Think Cool sessions and 68.46% of Act Cool sessions, while session-specific goals were rated 3 or 4 in 90.88% and 90.85% of sessions respectively. In addition, positive module differentiation was rated at the highest level in 90.39% of Think Cool sessions and 82.35% of Act Cool sessions. This indicates that not only was the intervention content delivered with good adherence, but also that the intended distinction between cognitive and behavioural components was largely preserved.

Wagner et al. (2014) also reported favourable fidelity findings, although with less detail than some of the studies above. All treatment sessions were audio-recorded, and discussed in weekly supervision, with independent research assistants rating 10% of session recordings on a 5-point scale. The mean fidelity score was 4.25, suggesting that the school-based CBT intervention Guided Self-Change was delivered with a high level of adherence. Relative to studies such as Esposito-Smythers et al. (2017) and Schwartz-Mette et al. (2024), the proportion of sessions formally rated was smaller and the reporting of specific fidelity dimensions was less extensive, but the use of independent raters and regular supervision still strengthens confidence in implementation quality.

Another shared feature across several studies was the use of training and supervision as safeguards against implementation drift. For instance, Lochman et al. (2017), described a structured fidelity framework for CP-IE. Detailed manuals were used for both child and parent sessions, and interventionists were master's-level clinicians who had previously completed multi-day training in the Coping Power model. Weekly supervision was provided by a doctoral-level psychologist, and co-leaders were also supervised regularly. Session checklists indicated that

objectives were fully covered in 91% of child sessions and 87% of parent sessions, with the remaining objectives partially covered (7% and 13%, respectively). Although these data were based on self-report rather than independent coding, they still suggest a high degree of procedural consistency. However, compared with studies that used independent observational methods, the evidence is somewhat less robust.

A similar pattern is visible in Daunic et al. (2006), Cole et al. (2013) and Watson (2009), where fidelity was generally high but supported by internal ratings and teacher or co-facilitator reports. In Daunic et al. (2006), teachers reported delivered lessons in the intended 30-minute format, covered all lesson content in 98% of lessons, and achieved strong participation in role-play activities (98%) and related practice exercises (91%). Project staff observations of 15 lessons broadly confirmed these reports, although one observed lesson omitted some review material and suggested resources. Cole et al. (2013) reported an overall mean fidelity score of 4.76 out of 5 (SD = 0.23) across six anger-management sessions, with no significant differences across facilitators or cohorts. Watson (2009) found a similar mean fidelity score of 4.77, with a range from 4.33 to 5. These findings indicate consistent adherence to programme structure, but the evidence is somewhat less persuasive than in studies where fidelity was assessed through more extensive independent coding.

The evidence from Cook et al. (2014) points to a slightly different implementation issue. Fidelity to the delivery of each intervention component was supported through manual use of BAM, use of research-hired staff for the Match tutoring model, and dosage records showing weekly BAM sessions and daily one-hour two-on-one maths tutoring. However, there were notable difficulties in preserving separation between treatment conditions. Young people in the BAM-only condition were sometimes sent to the Match classroom when disruptive, where tutors occasionally provided informal academic support. In addition, BAM groups included participants from both treatment arms, creating the possibility of peer-related spillover. This means that while intervention delivery itself may have followed the intended structure, fidelity to the distinct treatment arms. In this respect, the challenge lay not so much in therapist adherence as in maintaining clear boundaries between intervention components.

Augustyniak et al. (2009) and Pullen (1996) offer less robust fidelity evidence overall, though for different reasons. Augustyniak et al. (2009) described treatment integrity for the Prepare Curriculum as "excellent," supported by facilitator training, manual provision, and periodic follow-up by the authors. However, the absence of detailed quantitative adherence data makes this judgement harder to evaluate and less compelling than in studies reporting observed percentages or coded ratings. The authors also acknowledged that variation attributable to facilitator style and school context was not controlled for. Pullen (1996), by contrast, identified more direct concerns about implementation quality. Although all officers reported preparing for sessions, only 2 of 13 met the recommended two hours of preparation, while the remaining 11 officers averaged only 30 minutes. Seven officers began delivering the programme within the recommended two-month period after training, and three officers began after three months, but video-recorded sessions indicated that delivery often relied on reading from the manual rather than actively teaching the material. Compared with the other studies, this suggests weaker fidelity and points to limited preparation time and variable staff skill as implementation constraints.

Taken together, several common features appear to be associated with stronger fidelity across the studies: manualised content, clear training procedures, regular supervision, and direct review of recorded sessions. The evidence is most compelling where these features were paired with quantified adherence data and, ideally, independent coding, as seen in Esposito-Smythers et al. (2017), Schwartz-Mette et al. (2024) and Te Brinke et al. (2021). By contrast, confidence is more limited where fidelity was inferred primarily from self-report, broad statements of treatment integrity, or where contamination between conditions reduced clarity about what was delivered. Taken together, the studies suggest that most interventions were implemented with moderate to high fidelity, but they also indicate that the strength of this conclusion depends heavily on how fidelity was monitored and reported.

Reach and penetration

Reach and penetration refer to the extent to which the intervention has been integrated or reached eligible recipients. One very low quality study examined the reach and penetration of CBT Skills programmes. Pullen (1996) evaluated the

Reasoning and Rehabilitation cognitive skills development programme as part of the Juvenile Intensive Supervision Probation (JISP) through both an individual RCT and process evaluation. A total of 24 professionals were interviewed as part of the process evaluation, including thirteen probation officers, four JISP programme supervisors, three chief probation officers, two state Judicial Department administrators, and two programme trainers. The extent to which the intervention was integrated into the service setting and reached eligible participants appeared to vary and was influenced by how the programme was introduced and positioned within routine practice. While the programme was embedded within a juvenile justice context and offered to eligible young offenders, penetration was shaped by the degree to which practitioners actively promoted and normalised participation. Some officers reported routinely presenting the programme to all eligible individuals, suggesting an intention toward broad reach; however, the way in which it was framed often emphasised its mandatory nature rather than voluntary engagement. This may indicate that, although the intervention formally reached a proportion of the target population, true penetration in terms of meaningful uptake and engagement was less consistent. Adolescents were also perceived as more difficult to engage than adults, which may have further limited effective penetration among the intended population.

Efforts to enhance reach included preparatory strategies such as individual meetings with participants prior to programme commencement, aimed at building rapport, addressing concerns, and encouraging participation. In some cases, practitioners supplemented this with materials (e.g., informational flyers for youth and parents), and by framing the programme as beneficial or even enjoyable. However, variability in staff enthusiasm and approach to "selling" the programme suggests that integration into routine practice was uneven. Therefore, despite the intervention being available within the service, its reach and integration into routine practice were inconsistent and appeared to depend heavily on practitioner-led engagement efforts.

Sustainability

Sustainability refers to the ability of an implementer to maintain or institutionalise the intervention over time. No studies examined the sustainability of CBT Skills programmes.

Experiences of children and young people

Two studies examined the experiences of children and young people participating in CBT skills programmes, providing insight into how these interventions were perceived and experienced by participants. A high-quality cluster randomised controlled trial of the Coping Power – Internet Enhanced (CP-IE) programme (Lochman et al., 2017) and a lower-quality quasi-experimental study of an anger management CBT intervention (Watson, 2009) indicate that CBT skills programmes were generally experienced positively by children and young people, particularly when they supported emotional regulation, engagement, and peer relationships. However, the findings also highlight the importance of programme format, group dynamics, and delivery context in shaping participants' experiences.

Across both studies, children and young people reported that CBT programmes helped them to better understand and manage their emotions, particularly anger and behavioural difficulties. In Watson (2009), many children described anger as the primary reason for participation, with one child explaining that "sometimes I get really angry and just lose myself and like punch people," while another noted that "if any boy said anything, the slightest little thing, I get really angry." Children also described the negative consequences of their anger, including regret and getting into trouble, with one child stating, "I wished I didn't do anything now." Following participation in the programme, many reported improvements in emotional regulation and use of CBT strategies, such as "counting to twenty" to calm down, with participants noting that their "anger has gone down… lots" and that "after anger management I didn't react as much as I used to before the group." One child summarised the benefit by explaining that the programme "taught me to control my anger." Similarly, children in Lochman et al. (2017) rated the CP-IE programme highly overall (mean satisfaction score 6.63/7), suggesting that CBT-based skills training was perceived as helpful and relevant. The focus on goal setting, problem solving, and emotional regulation appeared to resonate with children and young people, supporting engagement and positive perceptions of the intervention.

Programme format and delivery were also important in shaping children and young people's experiences. In Lochman et al. (2017) participants reported that

in-person group sessions were particularly helpful (mean = 6.40), as they allowed for active discussion, practice of CBT skills, and interaction with peers in a structured school-based environment. The hybrid online component was also viewed positively, with children rating the website as helpful (mean = 6.08) and easy to access (mean = 5.70). Interactive features such as animated videos, quizzes, avatars, and reward systems appeared to enhance engagement and motivation, suggesting that combining digital and face-to-face delivery can support participation in CBT skills programmes. In contrast, children in Watson (2009) provided more varied feedback on programme delivery. Many of the children also described their enjoyment and how "fun" they found the group, suggesting engagement as a facilitator of positive experiences. While several children reported that they would not change anything because "it's just great," others suggested improvements such as "make it longer and then we can have a little bit of a play" and delivering the sessions in "a place that we like." These reflections highlight how practical aspects of delivery, including session length and environment, can influence children's engagement and overall experience of CBT programmes.

Peer relationships and group dynamics were particularly influential in shaping children's experiences in Watson (2009). Many participants emphasised the importance of trust and confidentiality, noting that being friends with others in the group meant "I know they're not going to tell anyone." Children also described supporting one another to use CBT strategies in everyday situations, such as "if I get stressed in football or something, they just say no don't cos you remember our group and I just get really happy and walk away." Some participants reported forming new friendships and improving relationships with peers, with one child explaining that "we didn't really like each other much, but now we do cos we've got to know each other through the group." However, not all experiences were positive, as a small number of children expressed concerns about confidentiality and trust, with one stating, "well I didn't mostly like the part where people who tell everybody about their private stuff." These findings suggest that group-based CBT programmes can foster supportive peer environments but require careful facilitation to ensure safety and mutual trust.

Overall, findings from both studies suggest that CBT Skills programmes were generally experienced positively by children and young people, particularly when

they supported emotional regulation, provided engaging and interactive delivery formats, and fostered supportive peer relationships. Watson (2009) qualitative data provides rich insight into how children understood and applied CBT skills in their everyday lives, while Lochman et al. (2017) demonstrates high levels of satisfaction with a structured hybrid intervention. However, the evidence base remains limited, as only two studies directly captured children's perspectives, and one relied primarily on quantitative satisfaction ratings rather than in-depth qualitative accounts. As a result, further research is needed to better understand how children and young people experience CBT Skills programmes across different contexts and delivery models, and how these programmes can be adapted to maximise engagement, trust, and perceived benefit.

How much does it cost?

One moderate quality study examined the cost of a CBT Skills programme. Cook et al. (2014) report the results of a RCT evaluating a two-pronged intervention designed to support disadvantaged youth who are falling behind in school through both academic and non-academic remediation. The intervention combines the Becoming a Man (BAM) programme, developed by Youth Guidance and grounded in principles of CBT, with intensive, individualised maths tutoring based on the Match Education model. The study involved 106 male students aged 14–16 in a public high school in Chicago, randomly assigned to receive either the BAM intervention alone, the combined BAM and tutoring intervention, or standard school supports.

The study includes a detailed financial analysis of the integrated intervention. The estimated cost per participant for delivering both the non-academic and academic components together is approximately $4,400, with a defensible range between $3,000 and $6,000, depending on implementation efficiency and scale. This variation reflects uncertainties of the tutoring component, particularly as tutor capacity was underutilised during implementation and it is unclear how this affected programme intensity.

The non-academic BAM component has a relatively stable and well-documented cost structure. It is estimated at approximately $1,900 per participant per academic year and is delivered in small to medium-sized groups of around 10 to

15 students. This estimate is based on sustained implementation experience by Youth Guidance and is therefore considered a reliable benchmark for this component of the intervention.

In contrast, the academic tutoring component presents greater cost variability. The authors note that their implementation operated at a "probably inefficiently small scale," resulting in excess supervisory capacity relative to the number of students served. As a result, their realised cost was approximately $4,000 per student. However, they estimate that if all programme slots had been filled, the cost would have been closer to $2,800 per student. To provide a more generalisable estimate, the authors adopt a benchmark cost of $2,500 per student based on large-scale implementation of the Match tutoring model reported by Fryer (2011). A key feature of this model is its reliance on instructional tutors rather than fully qualified teachers, allowing for significantly lower labour costs. Tutors are typically compensated at relatively modest levels according to the authors (approximately $16,000 plus benefits for the academic year), which makes the high-dosage tutoring model financially feasible.

In terms of cost-effectiveness, the authors conclude that the intervention delivers substantial improvements in outcomes per dollar spent compared to many prominent alternatives targeting disadvantaged youth. They compare the cost-adjusted impacts on maths test scores and overall grade point average (GPA) at the 10th-grade level with those from well-known interventions, including Perry Preschool Program, income support through the Earned Income Tax Credit (EITC), and class size reductions in early elementary education. Using present value cost estimates discounted to 10th grade, the authors find that their combined intervention performs favourably in terms of cost-effectiveness, including at the highest point of the cost range.

Overall, the analysis suggests that this integrated approach represents a cost-effective strategy for improving educational outcomes among disadvantaged adolescents. By combining evidence-based social-cognitive support with intensive tutoring delivered by lower-cost instructors, the intervention is able to generate large academic gains at a relatively moderate cost, positioning it as a promising and scalable cost-effective model. However, it is important to note that this cost analysis is based on a single study of one specific intervention

implemented in a particular context. As such, caution is warranted in generalising these findings across different settings, populations, and implementation conditions. Strengthening the evidence base is essential before drawing firm conclusions about the broader cost-effectiveness of CBT Skills interventions.

Conclusion and Takeaway Messages

Evidence from this systematic review and meta-analysis indicates that CBT Skills is effective in reducing violence, crime and offending, and a wider range of related outcomes amongst children and young people. Across 37 effectiveness studies involving 4,154 participants, CBT Skills appears to have a high impact on violence, a high impact on crime and offending, and a high overall impact across all measured outcomes. However, these findings should still be interpreted with some caution, as effect sizes varied substantially across studies and contexts, and the evidence base was much stronger for some outcomes and populations than for others.

Violence

CBT Skills interventions is estimated to have a high impact on violence, corresponding to an estimated 72% reduction in violence outcomes. This finding was based on 56 outcomes across 22 studies and had an evidence security rating of Level 3. The effect remained statistically significant after adjusting for clustering using robust variance estimation, suggesting that the overall violence finding is reasonably robust.

Importantly, violence in this review was defined broadly in line with YEF criteria and included not only violent crime and offending, but also aggressive and bullying-related behaviours directed at others. As such, this estimate captures meaningful reductions in violence-related behaviour more broadly, rather than only formal justice outcomes. In practical terms, this suggests that CBT Skills interventions can reduce aggression, bullying, conduct problems, and other behaviours closely linked to violence involvement amongst children and young people.

A consistent finding across multiple decades of research is that CBT-informed skill programmes reduce anger and aggression-related outcomes in youth, with effect sizes usually described as small to moderate overall but sometimes moderate to large in higher-risk groups and for outcomes closely tied to the taught skills (overall pooled d=0.67; skills specific training d=0.79; (Sukhodolsky et al., 2004).

Crime and Offending

CBT Skills interventions might have a high impact on crime and offending, corresponding to an estimated 71% reduction in crime and offending outcomes. This finding was based on 11 outcomes across four studies and remained statistically significant after robust variance estimation.

Although this is a very promising result, confidence in this finding is lower than for violence because it is based on a much smaller evidence base and has an evidence security rating of Level 1. Most of these outcomes focused on formal justice-related indicators such as recidivism, arrests, convictions, or technical violations, meaning the finding is especially important from a youth justice perspective. However, with only four studies contributing to this estimate, the size of the effect should be interpreted cautiously until it is replicated in a larger number of high-quality evaluations.

The wider evidence base provides some support for the proposition that CBT-based skills programmes reduce offending-related outcomes, however, much of this evidence comes from justice-involved samples rather than general school populations. A Campbell Collaboration systematic review on the effectiveness of CBT for offenders (adult and children) reported a reduction in average 12-month recidivism from 0.40 to 0.30, which translates to about a 25% relative decrease (Lipsey et al., 2007).

All outcomes

Looking across the full range of measured outcomes, CBT Skills interventions appears to lead to a high overall impact. The pooled effect across 251 outcomes from 37 studies was statistically significant and remained stable after adjustment for clustered outcomes. This indicates that CBT Skills interventions are not only likely to lead to reductions in violence and offending, but also likely to lead to broader improvements in outcomes linked to youth offending and violence.

Our moderator analyses suggest that CBT Skills interventions are particularly beneficial across behavioural difficulties, emotional regulation, school engagement, helping others, victimisation, and crime and offending outcomes. Overall, this suggests that CBT Skills interventions may improve both risk-related

behaviours and broader psychosocial functioning, which is important given that violence and offending are often shaped by multiple overlapping difficulties.

In the wider evidence base, on social skills training, these types of programmes typically show broad improvements spanning social-cognitive skills, prosocial behaviour, emotion regulation components, and as presented earlier, reductions in disruptive/conduct-related behaviours. Although the average magnitudes vary by setting, sample risk level, and outcome (Durlak et al., 2011). On example of this is a systematic meta-review (an umbrella type review) on youth violence prevention programmes which finds that some reported CBT-related effect sizes in prior meta-analyses falling in a broad 0.36–0.70 range across included reviews (Matjasko et al., 2012).

What Works

Evidence from thirteen studies indicates that CBT Skills programmes for children and young people are most effectively implemented when delivery is engaging, developmentally appropriate, feasible within school or service constraints, and supported by strong fidelity procedures. Effective implementation was typically enabled by practical, interactive activities; flexible or hybrid delivery formats (e.g., digital or group-based models); alignment with school routines; organisational buy-in; and manualised protocols with structured training and supervision. Feasibility improved when programmes reduced burden on staff and families, used shorter or adaptable sessions, incorporated varied activities, and involved teachers or parents. Fidelity was strengthened through systematic session monitoring and clear adherence procedures. Key barriers included organisational resistance, logistical challenges in schools, intensive demands on families, and variability in staff preparation and engagement.

Acceptability, appropriateness, feasibility, and adoption appeared closely interconnected across the evidence base. Programmes were more acceptable where they were engaging, practical, and manageable for children, families, and practitioners, particularly when interactive skill-building activities were relevant to everyday challenges and easy to integrate into routine practice. Appropriateness was strengthened where interventions were adapted to developmental needs and delivery contexts through digital components, brief classroom-friendly

sessions, and activity-based learning, while longer or more intensive programmes appeared less well matched to school or service environments. Feasibility was similarly shaped by contextual fit, with shorter and more flexible interventions easier to deliver within existing structures, whereas programmes requiring extensive coordination or family involvement were more vulnerable to logistical barriers. Adoption also depended on organisational readiness, with limited uptake in one study highlighting the importance of institutional support, early stakeholder engagement, and clear communication about programme feasibility and value.

High fidelity was another consistent feature of stronger implementation, supported by manualised protocols, structured training and supervision, and systematic monitoring of adherence through recorded sessions and fidelity checks. At the same time, reach and penetration, which was based on limited evidence, still appeared more variable, as the extent to which programmes were embedded in routine practice depended on practitioner engagement and local service capacity, suggesting that availability of CBT Skills programmes does not necessarily translate into consistent uptake across eligible children and young people.

However, most implementation evidence came from US and European school or service settings rather than UK contexts, which limits confidence in how far these findings transfer to routine UK practice. While one study did report cost-effectiveness evidence, economic data across the wider evidence base remain limited, and few studies provided detailed information on resource requirements, long-term costs, or scalability. As a result, firm conclusions about affordability, cost-effectiveness, and large-scale implementation of CBT Skills programmes cannot yet be drawn, highlighting the need for further research examining economic outcomes and sustainability within real-world educational and service systems.

Who Benefits Most?

The evidence suggests that CBT Skills interventions may be particularly effective for children and young people with externalising and behavioural dysregulation difficulties. This was the clearest and most robust subgroup finding and is likely

important for practice, given the close links between these difficulties and later violence or offending.

There was no evidence that effectiveness differed meaningfully by age group. Effects were remarkably similar across childhood, adolescence, and mixed-age samples, suggesting that CBT Skills interventions may be developmentally relevant across a broad age range when appropriately adapted.

Gender

Significant beneficial effects were observed in mixed-sex samples and in majority-male samples, while all-male samples showed a smaller, non-significant effect. However, the statistical evidence for differences between gender groups was only marginal, so firm conclusions should not be drawn.

Narrative evidence was similarly limited. Across the four studies that directly examined gender as a moderator of CBT Skills outcomes, three studies reported no significant gender differences across behavioural, emotional, or cognitive domains, indicating that boys and girls tended to benefit similarly from CBT-based programmes. Only one study identified gender-specific improvements, namely greater reductions in social-anxiety-related thoughts and verbal aggression among girls, but these effects were inconsistent across other outcomes and isolated within a single sample. Overall, the available evidence suggests broadly comparable effects across genders, with only tentative indication that certain emotional or aggression-related outcomes may vary.

As with the wider literature, however, the capacity to draw firm conclusions is severely constrained by structural data limitations. Many studies included too few female participants to detect meaningful differences, and where mixed-sex samples were included, gender analyses were often underpowered or absent. This mirrors the broader patterns identified in previous systematic reviews, such as Lipsey et al. (2007), where low female representation has hindered moderator analyses and limited understanding of intervention effects for girls. Consequently, while current findings tentatively suggest generally equitable outcomes for girls and boys, data remained underpowered and the evidence base remains too thin to determine with confidence whether CBT Skills interventions operate differently across genders. Addressing this gap will require future research that adequately includes and analytically prioritises female populations.

Ethnicity

A large proportion of studies (43%) did not report ethnicity at all, and several subgroup estimates were based on small numbers of studies. While some studies categorised as having diverse and Black and Global Majority samples showed promising effects, the pattern was not sufficiently clear or stable to support confident conclusions about differential effectiveness by ethnicity. This was mirrored in the narrative evidence. One study evaluating the Guided Self-Change (GSC) programme among a predominantly Black and Global Majority high school sample and found no evidence that ethnicity moderated treatment outcomes. Hispanic and African American students showed comparable reductions in alcohol use, drug use, and interpersonal aggression, suggesting broadly similar benefits across the ethnic groups included.

However, these analyses were limited to only two ethnic categories and did not capture broader racial, cultural, or linguistic diversity within the sample. The reliance on such aggregated groups, combined with limited representation across studies, means that the evidence base remains too narrow to assess whether CBT Skills interventions operate differently across a wider range of ethnic or cultural populations. As with gender, the inability to draw firm conclusions ultimately reflects structural data gaps across the field, underscoring the need for future research that includes more granular and diverse demographic data.

There were also major evidence gaps in relation to socioeconomic status, care experience, SEND, neurodiversity, place of residence, and intersectionality. Socioeconomic status was not reported in over half of the studies, only two studies provided narrative information relevant to care experience, and the wider evidence base does not allow any clear conclusions about how CBT Skills interventions work for these groups. This is a significant limitation, particularly given how over-represented many of these groups are within populations at risk of violence and offending.

The wider evidence base supports our finding that benefits are likely to occur across a range of demographic groups, but the clearest and most consistent "who benefits most?" signals relate less to age or gender and more to baseline risk, externalising difficulty, and implementation quality , while moderators like

gender and age show weaker and less consistent patterns (Beelmann & Lösel, 2006).

This pattern is echoed in broader school-based and justice-focused CBT reviews, which also find stronger effects where participants are at higher risk and where programmes are delivered with higher implementation quality (Lipsey et al., 2007). By contrast, the wider review literature provides much weaker and less consistent evidence that age, gender, or ethnicity reliably moderate outcomes, and recent equity-focused synthesis shows that most school-based violence prevention trials still do not report power, or test subgroup effects well enough to support firm conclusions for ethnicity, socioeconomic status, disability/SEND, or other marginalised groups (Eldred et al., 2026).

Limitations

Challenges with the studies measuring effectiveness

A key limitation of this evidence base is the very high heterogeneity across the main meta-analyses. This means that effect sizes varied substantially across studies, suggesting that differences in population, setting, intervention design, dosage, and outcome measurement all influence results. While the average effects are clearly positive, this level of variability means that effect sizes should not be assumed to transfer uniformly across contexts.

A second limitation is that findings on crime and offending were based on only four studies. This makes those estimates more vulnerable to instability and reduces confidence in their generalisability. Several subgroup analyses were also based on sparse categories, sometimes including only one or two studies, meaning apparently large effects in some subgroups should be interpreted very cautiously.

A third limitation is the inconsistency in demographic reporting. Gender, ethnicity, and socioeconomic status were often poorly reported, and information on care experience and other vulnerability-related characteristics was extremely limited. This significantly constrains our understanding of who benefits most from CBT Skills interventions.

A fourth limitation is that three otherwise eligible impact evaluations had to be excluded from the meta-analysis, which further reduced the available quantitative evidence. In each case, the studies met the inclusion criteria but did not report the statistical information needed to calculate a comparable between-group post-intervention effect size. Cole et al. (2013) lacked the raw descriptive statistics required for effect size extraction, Kozina (2021) did not provide a usable cluster-adjusted post-test estimate, and the raw data shared by the author were not accessible for use in this review, and Augustyniak et al. (2009) did not report sufficient post-intervention between-group data. All three studies were retained in the narrative synthesis, but their exclusion reduced the size and precision of the meta-analytic evidence base.

Finally, a further limitation concerns Saba et al. (2023), where the values reported were not clearly labelled. It was therefore unclear whether the variance represented standard deviations, standard errors, variances, or another metric. Treating them as standard deviations produced implausibly large effect sizes, so this interpretation was not considered sufficiently secure for the primary analysis. Sensitivity analyses were therefore conducted using alternative assumptions, and the more plausible estimates were retained in the main analysis (see Appendix 3 for a further description). However, because the scale of these values could not be verified from the paper, findings from this study should be interpreted with caution.

Challenges with the studies measuring implementation

Thirteen studies were eligible for inclusion; however, coverage of implementation outcomes was uneven. Evidence was identified for acceptability, appropriateness, adoption, feasibility, fidelity and reach/penetration. Notably, adoption and reach were each examined by only one low or very low-quality study, and no studies meaningfully examined sustainability. This limits our understanding of how CBT Skills programmes are taken up across services, embedded into routine systems, and maintained over time.

Most included studies were quantitative and were primarily designed to test intervention effects rather than implementation. As a result, implementation outcomes were often reported as secondary or incidental findings rather than examined in a systematic way. Fidelity was the clearest example of this, as it was

often included to protect internal validity within trials through manuals, training, supervision, and session checks.

In addition, the way implementation outcomes were measured varied substantially across studies. Some outcomes were assessed directly through satisfaction ratings, observations, recordings, teachers' perceptions or adherence checklists, but many were inferred through more indirect indicators such as attendance, retention, treatment completion, or authors' reflections on programme fit. In several studies, judgements about implementation relied heavily on self-report from facilitators, teachers, or practitioners, which makes it harder to judge how accurately these findings reflect delivery in practice. This inconsistency in measurement also makes it difficult to compare findings across studies or draw firm conclusions about which implementation approaches are most effective.

There was also limited evidence about implementation under routine service conditions. Many studies were carried out in research-led, closely supported environments where delivery was carefully monitored, and staff often received more substantial training and supervision than is often feasible in routine services. While this may strengthen confidence that interventions were delivered as intended during the study period, it may also overestimate how easily the same programmes could be implemented in everyday school, community, or service settings where time, staffing, and resources are more constrained.

Finally, although cost and resource reporting was largely absent across the evidence base, one study did provide useful economic detail (Cook et al. 2014). Although the authors concluded that the intervention compared favourably with several other youth-focused programmes on a cost-adjusted basis, this evidence came from a single study in one specific context. As such, the broader evidence base remains limited in its ability to judge affordability, resource demands, or cost-effectiveness of implementing CBT Skills programmes at scale.

Final Thoughts and Recommendations

Overall, CBT Skills interventions appear to be a highly promising approach for reducing violence, crime and offending amongst children and young people. The evidence suggests high impact on violence, high impact on crime and offending,

and broader positive effects across a range of related outcomes. The strongest and most reliable evidence currently comes from school-based delivery, and from samples of children and young people with externalising or behavioural dysregulation difficulties. There is also some indication that CBT Skills interventions can be effective across developmental stages and across a range of delivery formats and intensities.

At the same time, caution is needed. The evidence base is highly heterogeneous, crime and offending findings are underpinned by only a small number of studies, and subgroup findings are often limited by sparse data and inconsistent reporting. As such, CBT Skills interventions should be viewed as a strong and promising approach, but one whose effects are likely to depend on who receives them, where they are delivered, and how well they are implemented.

Based on the current available evidence, we make the following key recommendations:

- CBT Skills interventions should be prioritised as targeted interventions for children and young people with externalising, behavioural dysregulation, aggression, or related risk profiles, given that this is where the clearest evidence of benefit currently lies.
- Given the strong school-based evidence, schools and education-linked services should be viewed as a key setting for delivery, while further research is undertaken in clinical, residential, justice, and digital settings.
- Services should not assume that higher intensity automatically leads to better outcomes. Brief and moderate-intensity CBT Skills interventions may be capable of achieving meaningful benefits and may offer a more scalable option in line with the implementation evidence presented.
- More high-quality studies are needed on crime and offending outcomes specifically, including studies using administrative or official offending measures, longer follow-up periods, and larger samples.
- Future evaluations should report demographic and vulnerability-related characteristics more consistently, including ethnicity, socioeconomic status, care experience, SEND, neurodiversity, and intersectional identities, so that the field can better understand who benefits most.

- Given the breadth of positive outcomes observed, CBT Skills interventions should be understood not only as a violence reduction approach, but also as interventions that can improve emotional regulation, behaviour, school engagement, and related factors that shape longer-term life chances.
- Design CBT Skills programmes so they can be delivered within routine practice, with realistic session length, scheduling, and staffing demands for schools and services.
- Adapt interventions to children and young people's developmental level by using clear, concrete and engaging materials while keeping core CBT content intact. Co-design approaches involving young people, families, and practitioners may help ensure materials are developmentally appropriate and meaningful for end users.
- Build in regular feedback from children and young people, alongside parent and practitioner views, to ensure implementation decisions are informed by how acceptable and manageable programmes are in practice.
- Implementers and researchers should routinely document the resources required for CBT Skills programme delivery such as staffing, training, supervision, and materials. Clear reporting of costs is essential for assessing affordability, cost-effectiveness, and the feasibility of wider implementation.

References

Studies included in effectiveness and implementation sections are marked with an asterisk (*) at the beginning of the reference.

* Augustyniak, K. M., Brooks, M., Rinaldo, V. J., Bogner, R., & Hodges, S. (2009). Emotional Regulation: Considerations for School-Based Group Interventions. The Journal for Specialists in Group Work, 34(4), 326–350. https://doi.org/10.1080/01933920903219060

* Avci, D., & Kelleci, M. (2016). Effects of the Anger Coping Programme based on cognitive behavioural techniques on adolescents' anger, aggression and psychological symptoms. International Journal of Nursing Practice, 22(2), 189–196. https://doi.org/10.1111/ijn.12410

Barrett, H., Slesnick, N., Brody, J. L., Turner, C. W., & Peterson, T. R. (2001). Treatment outcomes for adolescent substance abuse at 4- and 7-month assessments. Journal of Consulting and Clinical Psychology, 69(5), 802–813. https://doi.org/10.1037/0022-006X.69.5.802

Beelmann, A., & Lösel, F. (2006). Child social skills training in developmental crime prevention: Effects on antisocial behavior and social competence. Psicothema, 18(3), 603–610.

* Bowman, P. C., & Auerbach, S. M. (1982). Impulsive Youthful Offenders: A Multimodal Cognitive Behavioral Treatment Program. Criminal Justice and Behavior, 9(4), 432–454. https://doi.org/10.1177/0093854882009004003

* Burraston, B. O., Cherrington, D. J., & Bahr, S. J. (2012). Reducing Juvenile Recidivism With Cognitive Training and a Cell Phone Follow-Up: An Evaluation of the RealVictory Program. International Journal of Offender Therapy and Comparative Criminology, 56(1), 61–80. https://doi.org/10.1177/0306624X10388635

* Cole, R. L., Treadwell, S., Dosani, S., & Frederickson, N. (2013). Evaluation of a short-term, cognitive-behavioral intervention for primary age children with anger-related difficulties. School Psychology International, 34(1), 82–100. https://doi.org/10.1177/0143034312451062

* Cook, P., Dodge, K., Farkas, G., Fryer, R., Guryan, J., Ludwig, J., Mayer, S., Pollack, H., & Steinberg, L. (2014). The (Surprising) Efficacy of Academic and Behavioral

Intervention with Disadvantaged Youth: Results from a Randomized Experiment in Chicago (No. W19862; p. w19862). National Bureau of Economic Research. https://doi.org/10.3386/w19862

* Daunic, A. P., Smith, S. W., Brank, E. M., & Penfield, R. D. (2006). Classroom-based cognitive–behavioral intervention to prevent aggression: Efficacy and social validity. Journal of School Psychology, 44(2), 123–139. https://doi.org/10.1016/j.jsp.2006.01.005

* Doğan, S., & Çam, O. (2020). The effect of adaptation theory-based therapeutic approach on adolescents anger management. Journal of Child and Adolescent Psychiatric Nursing, 33(2), 85–94. https://doi.org/10.1111/jcap.12268

Durlak, J. A., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. Child Development, 82(1), 405–432. https://doi.org/10.1111/j.1467-8624.2010.01564.x

Eldred, E., Devries, K., Nelson, K. A., Zinke-Allmang, A., Opondo, C., Mughis, W., Mallick, R., Banks, L. M., & Bhatia, A. (2026). A systematic review and equity analysis of school-based violence prevention interventions evaluated in randomised controlled trials. BMC Public Health, 26(1), 681. https://doi.org/10.1186/s12889-025-26142-1

* Esposito-Smythers, C., Hadley, W., Curby, T. W., & Brown, L. K. (2017). Randomized pilot trial of a cognitive-behavioral alcohol, self-harm, and HIV prevention program for teens in mental health treatment. Behaviour Research and Therapy, 89, 49–56. https://doi.org/10.1016/j.brat.2016.11.005

* Etscheidt, S. (1991). Reducing Aggressive Behavior and Improving Self-Control: A Cognitive-Behavioral Training Program for Behaviorally Disordered Adolescents. Behavioral Disorders, 16(2), 107–115. https://doi.org/10.1177/019874299101600204

* Freiden, J. (2006). GAME: A clinical intervention to reduce adolescent violence in schools [The University of Memphis]. https://www.proquest.com/openview/a320075fb86657ce3d721815a9560715/1?cbl=18750&diss=y&pq-origsite=gscholar

* Glick, B., & Goldstein, A. P. (1987). Aggression Replacement Training. Journal of Counseling & Development, 65(7), 356–362. https://doi.org/10.1002/j.1556-6676.1987.tb00730.x

* Hawkins, J. D., Jenson, J. M., Catalano, R. F., & Wells, E. A. (1991). Effects of a Skills Training Intervention With Juvenile Delinquents. Research on Social Work Practice, 1(2), 107–121. https://doi.org/10.1177/104973159100100201

Hoogsteder, L. M., Stams, G.-J. J. M., Schippers, E. E., & Bonnes, D. (2018). Responsive Aggression Regulation Therapy (Re-ART): An Evaluation Study in a Dutch Juvenile Justice Institution in Terms of Recidivism. International Journal of Offender Therapy and Comparative Criminology, 62(14), 4403–4424. https://doi.org/10.1177/0306624X18761267

* Ime, Y. (2024). The Effect of Cognitive Behavioral Group Counseling on Bullying and Empathy Levels of Adolescents. Anales de Psicología, 41(1), 85–93. https://doi.org/10.6018/analesps.601671

* Johnsen, D. B., Lomholt, J. J., Heyne, D., Jensen, M. B., Jeppesen, P., Silverman, W. K., & Thastum, M. (2024). The Effectiveness of Modular Transdiagnostic Cognitive Behavioral Therapy Versus Treatment as Usual for Youths Displaying School Attendance Problems: A Randomized Controlled Trial. Research on Child and Adolescent Psychopathology, 52(9), 1397–1412. https://doi.org/10.1007/s10802-024-01196-8

* Karatas, Z., & Gokcakan, Z. (2009). A Comparative Investigation of the Effects of Cognitive-Behavioral Group Practices and Psychodrama on adolescent Aggression. 2009. https://eric.ed.gov/?id=EJ858929

* Kettlewell, P. W., & Kausch, D. F. (1983). The generalization of the effects of a cognitive-behavioral treatment program for aggressive children. Journal of Abnormal Child Psychology, 11(1), 101–114. https://doi.org/10.1007/BF00912181

Kozina, A. (2021). Can FRIENDS for Life social-emotional learning programme be used for preventing anxiety and aggression in a school environment: 6 months, 1-year and 1-and-a-half-year follow-up. European Journal of Developmental Psychology, 18(2), 214–229. https://doi.org/10.1080/17405629.2020.1776103

* Kumuyi, D. O., Akinnawo, E. O., Akpunne, B. C., Akintola, A. A., Onisile, D. F., & Aniemeka, O. O. (2022). Effectiveness of cognitive behavioural therapy and social

skills training in management of conduct disorder. South African Journal of Psychiatry, 28. https://doi.org/10.4102/sajpsychiatry.v28i0.1737

* Lesure-Lester, G. E. (2002). An Application of Cognitive-Behavior Principles in the Reduction of Aggression Among Abused African American Adolescents. Journal of Interpersonal Violence, 17(4), 394–402. https://doi.org/10.1177/0886260502017004003

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

* Lochman, J. E., Boxmeyer, C. L., Jones, S., Qu, L., Ewoldsen, D., & Nelson, W. M. (2017). Testing the feasibility of a briefer school-based preventive intervention with aggressive children: A hybrid intervention with face-to-face and internet components. Journal of School Psychology, 62, 33–50. https://doi.org/10.1016/j.jsp.2017.03.010

* Logsdon, C. D. (2003). Early intervention with aggressive boys: A cognitive-behavioral group intervention in a school setting [State University of New York at Albany]. https://www.proquest.com/openview/2218f18fbf884fa01f593b0b7e8f5532/1?pq-origsite=gscholar&cbl=18750&diss=y

Matjasko, J. L., Vivolo-Kantor, A. M., Massetti, G. M., Holland, K. M., Holt, M. K., & Dela Cruz, J. (2012). A systematic meta-review of evaluations of youth violence prevention programs: Common and divergent findings from 25years of meta-analyses and systematic reviews. Aggression and Violent Behavior, 17(6), 540–552. https://doi.org/10.1016/j.avb.2012.06.006

* McLaughlin, L. P. (2009). The effect of cognitive behavioral therapy and cognitive behavioral therapy plus media on the reduction of bullying and victimization and the increase of empathy and bystander response in a bully prevention program for urban sixth-grade students [The University of Toledo]. https://www.proquest.com/openview/95b01029f9258aed8708cfe66835c34b/1?pq-origsite=gscholar&cbl=18750

* Moynahan, L., & Strømgren, B. (2005). Preliminary results of Aggression Replacement Training for Norwegian youth with aggressive behaviour and with a

different diagnosis. Psychology, Crime & Law, 11(4), 411–419. https://doi.org/10.1080/10683160500256784

* Njardvik, U., Smaradottir, H., & Öst, L.-G. (2022). The Effects of Emotion Regulation Treatment on Disruptive Behavior Problems in Children: A Randomized Controlled Trial. Research on Child and Adolescent Psychopathology, 50(7), 895–905. https://doi.org/10.1007/s10802-022-00903-7

* Nwolisa, F. A. (2023). Assessment and management of psychosocial problems of adolescents in correctional centres in Lagos state, Nigeria.

* Oparaduru, J. O. (2017). EFFECTIVENESS OF COGNITIVE RESTRUCTURING AND SELF-CONTROL ON PRONENESS TO MALADAPTIVE BEHAVIOUR AMONG SELECTED SENIOR SECONDARY SCHOOL STUDENTS IN IMO STATE, NIGERIA.E. Https://Ir.unilag.edu.ng/Handle/123456789/3266

* Orgilés, M., Morales, A., Fernández-Martínez, I., Méndez, X., & Espada, J. P. (2023). Effectiveness of a transdiagnostic computerized self-applied program targeting children with emotional problems: A randomized controlled trial. Journal of Affective Disorders, 338, 155–162. https://doi.org/10.1016/j.jad.2023.06.004

Parker, K., Nunns, M., Xiao, Z., Ford, T., Stallard, P., Kuyken, W., Axford, N., & Ukoumunne, O. C. (2025). Patterns of intra-cluster correlation coefficients in school-based cluster randomised controlled trials of interventions for improving social-emotional functioning outcomes in pupils: A secondary data analysis of five UK-based studies. BMC Medical Research Methodology, 25(1), 120. https://doi.org/10.1186/s12874-025-02574-6

* Poulin, F., Dishion, T. J., & Burraston, B. (2001). 3-Year Iatrogenic Effects Associated With Aggregating High-Risk Adolescents in Cognitive-Behavioral Preventive Interventions. Applied Developmental Science, 5(4), 214–224. https://doi.org/10.1207/S1532480XADS0504_03

* Pullen, S. (1996). Evaluation of the Reasoning and Rehabilitation cognitive skills development program as implemented in juvenile ISP in Colorado.

Pustejovsky, J. E., & Tipton, E. (2022). Meta-analysis with Robust Variance Estimation: Expanding the Range of Working Models. Prevention Science, 23(3), 425–438. https://doi.org/10.1007/s11121-021-01246-3

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. Psychological Bulletin, 86(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638

Rutterford, C., Copas, A., & Eldridge, S. (2015). Methods for sample size determination in cluster randomized trials. International Journal of Epidemiology, 44(3), 1051–1067. https://doi.org/10.1093/ije/dyv113

* Saba, P., Qi, H., Saleem, A., Chen, I.-J., Kausar, F. N., & Iqbal, M. F. (2023). Effects of Animated Movies on the Aggression and Behavior Performance of Primary School Students and Their Control Using a Cognitive Behavioral Anger-Control Training (CBACT) Program. Behavioral Sciences, 13(8), 659. https://doi.org/10.3390/bs13080659

* Schwartz-Mette, R. A., Lawrence, H. R., Fearey, E., Shankman, J., Nichols, J., Walters, J., Perello, E., & Smith, S. (2024). A School-Based Evaluation of the FRIENDS Resilience Programs: Implications for Mental Health Concerns in Rural Students. School Mental Health, 16(4), 1156–1170. https://doi.org/10.1007/s12310-024-09682-6

* Squires, G., & Caddick, K. (2012). Using group cognitive behavioural therapy intervention in school settings with pupils who have externalizing behavioural difficulties: An unexpected result. Emotional and Behavioural Difficulties, 17(1), 25–45. https://doi.org/10.1080/13632752.2012.652423

Sukhodolsky, D. G., Kassinove, H., & Gorman, B. S. (2004). Cognitive-behavioral therapy for anger in children and adolescents: A meta-analysis. Aggression and Violent Behavior, 9(3), 247–269. https://doi.org/10.1016/j.avb.2003.08.005

* Sukhodolsky, D. G., Solomon, R. M., & Perine, J. (2000). Cognitive-Behavioral, Anger-Control Intervention for Elementary School Children: A Treatment-Outcome Study. Journal of Child and Adolescent Group Therapy, 10(3), 159–170. https://doi.org/10.1023/A:1009488701736

* Te Brinke, L. W., Menting, A. T. A., Schuiringa, H. D., Deković, M., Weisz, J. R., & De Castro, B. O. (2021). Emotion regulation training as a treatment element for externalizing problems in adolescence: A randomized controlled micro-trial. Behaviour Research and Therapy, 143, 103889. https://doi.org/10.1016/j.brat.2021.103889

* Wagner, E. F., Hospital, M. M., Graziano, J. N., Morris, S. L., & Gil, A. G. (2014). A randomized controlled trial of guided self-change with minority adolescents.

Journal of Consulting and Clinical Psychology, 82(6), 1128–1139. https://doi.org/10.1037/a0036939

* Watson, S. E. (2009). An evaluation of a short term cognitive behavioural anger management intervention for key stage two pupils [Institute of Education, University of London]. https://discovery.ucl.ac.uk/id/eprint/10007495/

YEF. (2023). Children, violence and vulnerability: The second annual Youth Endowment Fund report into young people's experiences of violence. Youth Endowment Fund. https://youthendowmentfund.org.uk/wp-content/uploads/2023/11/YEF_Children_violence_and_vulnerability_2023_FINAL.pdf

* Yeo, L. S., & Choi, P. M. (2011). Cognitive-behavioural therapy for children with behavioural difficulties in the Singapore mainstream school setting. School Psychology International, 32(6), 616–631. https://doi.org/10.1177/0143034311406820

Appendix 1. Methods of the systematic review

Protocol

Prior to initiating this systematic review, we developed a comprehensive protocol for an Evidence and Gap Map (EGM) outlining the research objectives, eligibility criteria, search strategy, data extraction, quality appraisal, and synthesis methods. This protocol was registered and is available on the Open Science Framework (OSF),[6] ensuring transparency and adherence to predefined methods.

Eligibility Criteria

To define the scope of relevant research, we applied the following criteria from the outset:

Table 21: *Eligibility criteria for CBT Skills studies*

Eligibility criteria	Inclusion criteria	Exclusion criteria
Population	Children and young people aged 0-17. General population; Children and young people exhibiting behavioural difficulties or at risk of or have offended, but typically without a requirement for a formal clinical diagnosis.	Studies with samples predominantly composed of adults aged 18 years and older Where studies include wider age bands, data must be extractable for participants aged 0–17 (or the sample mean age must be ≤17 and the vast majority of participants under 18
Intervention	Programmes using CBT techniques or principles (e.g.,	Interventions that are purely educational or mentoring without a specific CBT

[6] Protocol is available to access here: https://osf.io/vamxy

	anger management, social problem-solving). Interventions must explicitly state the use of CBT elements.	component; other therapeutic interventions without CBT element; pharmacological interventions; clinical CBT requiring diagnosis or delivered primarily for mental health treatment; Multi-component interventions where CBT is minor (<30% of time) Trauma-focused therapies delivered primarily for victims (e.g., PTSD-focused TF-CBT victim recovery model), where the theory of change's primary aim is addressing PTSD/trauma symptoms rather than violence prevention in at-risk/offending populations.
Comparator	Business as usual / treatment as usual (BAU/TAU), no intervention, waitlist, or minimal-contact control In terms of study design, we only include evaluations with a credible counterfactual (i.e. an RCT or a robust quasi-experimental design such as DiD, synthetic control, regression discontinuity, interrupted time series with control, instrumental variables, or matched comparison).	We will exclude evaluations where the comparator is another structured active psychosocial intervention (e.g., another named therapy and/ or curriculum) unless it clearly constitutes BAU within that service context

Dose	Low to Moderate**.** Typically fixed duration (e.g., 10-20 sessions) or <30 contact hours. Focused on skills transfer. Programmes should devote ≥30–50% of contact time to CBT	High-intensity, individualised, and tailored CBT programmes where the CBT component comprises <30% of contact time. Programmes with modules mainly on psychoeducation about traumatic stress, triggers, affect regulation, trauma narrative/processing, grief/loss processing, trauma-related cognitions
Setting	Any non-clinical setting, e.g., schools, community centres, youth clubs, youth justice settings, home-based	Clinical settings: Hospitals, inpatient psychiatric units, or routine clinical care pathways (CAMHS).
Activities	Group or individual sessions focused on behavioural change via cognitive techniques.	Programmes with modules mainly on psychoeducation about traumatic stress, triggers, affect regulation, trauma narrative/processing, grief/loss processing, trauma-related cognitions
Participation	Universal (all youth in a setting), selected (at-risk), or indicated (already offending).	na
Who delivers the intervention	Non-clinical staff (teachers, youth workers, probation officers, mentors, peers) OR clinical staff delivering in a non-clinical 'programme' format.	na

Timing	Short term	Long-term, continuous care models exceeding 6 months
Study design	Randomised controlled trials (RCTs) and robust quasi-experimental designs (QEDs) e.g. those using difference-in-difference, synthetic controls, regression discontinuity, propensity score matching, or instrumental variables.	Studies without a credible counterfactual e.g. pre/post designs, QEDs with non-robust matching techniques

Details of searching

Prior to running academic database searches, the review team had already identified a number of potentially relevant studies from existing sources (n=121). These included studies coded within the Evidence and Gap Map by the previous team, studies already included at the EGM stage by the current team, and references identified from previous reviews. In addition, a further set of records was identified through academic database searching.

The academic search strategy was developed iteratively and refined through testing in MEDLINE. A set of pre-identified benchmark studies were identified by YEF at scoping stage, selected for their relevance to the review parameters and their coverage of key populations, interventions, and outcomes, was used to assess the sensitivity of the search. The strategy was refined until all benchmark studies were successfully retrieved, providing confidence that the search was capable of identifying relevant literature. The search was designed to capture both CBT Skills programmes and Clinical CBT.

The final MEDLINE search strategy was subsequently adapted, with appropriate modifications to controlled vocabulary, syntax, and field codes, for use across additional databases and platforms.

Searches were conducted in the following sources:

- Ovid MEDLINE (ALL)
- Child Development & Adolescent Studies (EBSCOhost)

- Open Dissertations (EBSCOhost)
- Criminal Justice Abstracts (EBSCOhost)
- ERIC (EBSCOhost)
- APA PsycInfo (Ovid)
- PsycArticles (Ovid)
- Embase (Ovid)

These databases were selected to ensure comprehensive coverage across psychology, education, criminal justice, health, and grey literature sources. The core conceptual structure of the search was retained across platforms, with database-specific adaptations applied to maximise sensitivity while maintaining relevance.

Details of screening

A total of 11,500 records were identified as potentially relevant for the CBT strands. These records were first deduplicated in EndNote, leaving 9,726 records to be imported into EPPI-Reviewer. Following further deduplication in EPPI-Reviewer, 8,651 records remained for title and abstract screening.

The EPPI-Reviewer classifier conducted initial title and abstract screening. A total of 1,772 records were marked as included and 6,879 as excluded. As a quality assurance check, a random sample of 10% of records excluded by the classifier was reviewed by senior team members.

A team of reviewers then screened the full text for all included studies. Senior reviewers conducted a check of any exclusions, and any discrepancies were discussed as a team and reconciled.

Of the 1,772 records screened at full-text level, 1,002 were excluded (for reasons, see Table 22). For inaccessible PDFs, the team attempted to contact lead authors to request access to the report or further data.

Table 22: *Full-text screening results*

Reason for exclusion	Number of Records Excluded at Full-Text Level
Study design not meeting robust causal threshold	343

Outcomes or intervention not relevant	545
Did not target CYP	82
Better placed in a different YEF Toolkit strand	5
Language	31
Duplicate	1

Following full-text screening, 770 studies were flagged as potentially relevant for inclusion. When combined with the 121 records identified via other sources, this gave a total of 891 records assessed as potentially relevant for inclusion.

Table 23: **Reasons for exclusion after full text screening**

Reason for exclusion	Number of Records Excluded
Retained as EGM records but not causal or CBT-relevant	863
Study design	2
Study belongs in another Toolkit strand	1
Outcomes not disaggregated	1

Initially the team identified 42 CBT Skills interventions which all went through our normal process for inclusion based on our scoping note. Following this final assessment by senior team members, 38 records were finally identified as fitting the scope for CBT Skills.

Of the 38 unique study records, 13 studies provided implementation evidence, and 38 studies provided effectiveness evidence. Of these 38 studies, 3 did not provide usable quantitative data for meta-analysis. The reasons for excluding these three records from the meta-analysis were as follows:

- **Cole et al. (2013)** met the inclusion criteria for an impact evaluation, using a matched, cluster-randomised waitlist control design. However, the study could not be included in the final meta-analysis due to data missingness.

The authors reported omnibus repeated-measures F-values and within-group post-hoc tests, but did not report the raw descriptive statistics required for effect size extraction. The study authors were contacted to request the missing raw data, but no response was received. As this essential statistical information was not extractable and could not be obtained from the authors, the study was excluded from the meta-analysis but was retained in the narrative synthesis of the effectiveness and implementation evidence.

- **Kozina (2021)** also met the inclusion criteria for an impact evaluation. The study used a cluster-randomised by class design, with four classes in total, and outcomes were analysed using repeated-measures mixed ANOVA across five time points, with gender included as an additional factor. However, the published statistics did not provide a post-only, cluster-adjusted between-group effect size suitable for meta-analysis. The author was contacted and provided all their raw data, but these were not in a format that could not be accessed and used for the present review. The study was therefore excluded from the meta-analysis but was considered in the narrative description of the effectiveness evidence.
- **Augustyniak et al. (2009)** met the inclusion criteria for an impact evaluation but was excluded from the quantitative synthesis because the paper did not report sufficient between-group post-intervention data to calculate an effect size for treatment versus control. Although the study included treatment and control groups assessed at pre-test and post-test, the results were reported primarily as paired-samples t-tests for within-group change in the treatment group, alongside narrative statements about post-treatment comparisons. The paper did not provide group-specific post-test means and standard deviations, or another convertible between-group effect estimate, for the relevant outcomes. The study was therefore excluded from the meta-analysis but retained for narrative synthesis of the effectiveness and implementation evidence.

Therefore, this report includes **38 unique studies**:

- **13 studies** which provided implementation evidence,
- **38 studies** which provided effectiveness evidence in the broader review,

- 35 studies, including two multi-arm trials that contributed independent comparisons, provided usable quantitative data for the meta-analysis, resulting in **37 effectiveness study** contributions overall.

All exclusions made following full-text screening were reviewed by a senior reviewer to ensure consistency and accuracy.

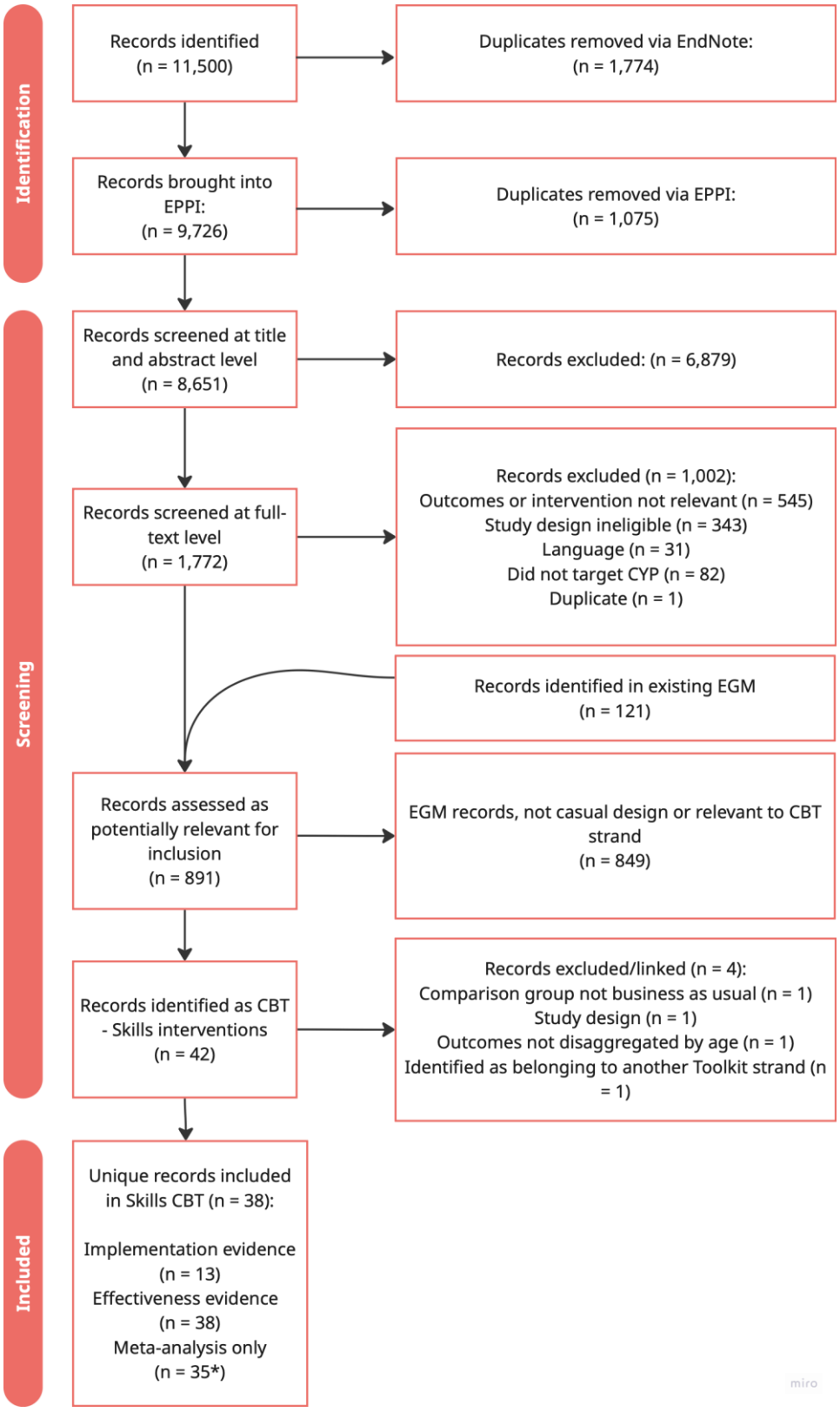

Figure 7: *PRISMA Flow Diagram*

Of the 35 included studies, 2 included two multi-arm trials that contributed independent comparisons for inclusion in the meta-analysis, resulting in 37 effectiveness study contributions overall

Quality appraisal process for effectiveness studies

The YEF-EQA tool was used across all studies to systematically assess the quality, reliability, and relevance of the research.

Table 24: *Quality appraisal ratings for studies included in the CBT Skills strand*

Study ID	Overall quality of the study	Study Design
Augustyniak (2009)	Low	RCT and PE
Avci (2016)	Moderate	RCT
Bowman (1982)	Low	RCT
Burraston (2012)	Moderate	QED
Cole et al. (2013)	Moderate	RCT and PE
Cook (2014)	Moderate	RCT
Daunic (2006)	High	RCT and PE
Dogan (2020)	Moderate	RCT
Esposito-Smythers (2017)	Moderate	RCT and PE
Etscheidt (1991)	Moderate	RCT
Freiden (2006)	Low	RCT
Glick (1987)	Low	RCT
Hawkins (1991)	Moderate	RCT
Ime (2025)	Moderate	RCT

Johnsen (2024)	Moderate	RCT
Karatas (2009)	Low	RCT
Kettlewell (1983)	Moderate	RCT
Kumuyi (2022)	Moderate	RCT
LeSure-Lester (2002)	High	RCT
Lochman (2017)	High	RCT and PE
Logsdon (2004)	Low	RCT and PE
McLaughlin (2009)	Moderate	RCT
Moynahan (2005)	Low	RCT
Njardvik (2022)	Moderate	RCT
Nwolisa (2023)	Moderate	RCT
Oparaduru (2017)	Moderate	QED
Orgilés (2023)	Moderate	RCT and PE
Poulin (2001)	Moderate	QED and PE
Pullen (1996)	Very low	RCT and PE
Saba (2023)	Low	RCT
Schwartz-Mette (2024)	Moderate	RCT and PE
Squires (2012)	High	RCT
Sukhodolsky (2000)	Moderate	RCT and PE
Te Brinke (2021)	Moderate	RCT and PE
Wagner (2014)	Moderate	RCT
Watson (2009)	Low	QED

Yeo (2011)	High	RCT

How the findings were analysed and combined

Preparing the data frame for analysis

Included studies comprised a mixture of individually randomised controlled trials, cluster randomised controlled trials, and robust quasi-experimental designs. For all studies, the primary effect size was the standardised mean difference (SMD) expressed as Hedges' g.

As is common in meta-analysis, the team encountered several challenges in harmonising effect sizes across studies. First, there was a need to standardise the direction of SMD values so that negative values consistently indicate a favourable intervention effect. To achieve alignment across all outcomes, we reviewed all effect size direction labels and, where necessary, multiplied positive SMD values by –1 for outcomes where a lower score signified improvement This ensured that all SMDs reflected the same directional meaning, **that negative SMD values always indicate beneficial effects.** This approach supports clear interpretation and comparability across studies within the meta-analysis.

Two outcomes were odds ratios and so were approximately converted into an equivalent SMD by multiplying the natural log of the OR $\times \frac{\sqrt{3}}{\pi}$ (approximately 0.5513).[7] In practical terms:

$$SMD\ (Cohen's\ d)\ \approx \frac{\sqrt{3}}{\pi}\ln(OR)\ =\ \frac{\ln(OR)}{1.814}$$

Additionally, eight outcomes across two studies were from cluster randomised controlled trials. In a cluster trial, randomisation occurs at a group-level (e.g. school, classroom). People within the same cluster tend to be more similar to each other than to people in other clusters. There is therefore less information about the effectiveness of an intervention in a trial of 100 people in 10 schools than

[7] Because this transformation is a linear scaling, the standard error of the SMD can be obtained by applying the same $\frac{\sqrt{3}}{\pi}$ factor to the standard error of the log OR

in a trial of 100 people randomised individually. This non-independence of participants within clusters must be accounted for or the effect size will be over-estimated. For cluster randomised trials which reported the effect sizes from cluster-adjusted analyses (e.g. mixed-effects models, generalised estimating equations) along with their standard errors or confidence intervals, we converted those estimates directly to Hedges' g and derived the standard error of g from the reported uncertainty, with no further adjustment required. However, for cluster randomised trials analysed as if they were individually randomised, the effect size precision was adjusted using a design effect (Rutterford et al., 2015). The design effect (DE) was defined as:

$$DE = 1 + (M - 1)\rho$$

Where M is the average cluster size and ρ is the intracluster correlation coefficient (ICC). For these studies we retained the point estimate of g but inflated its standard error by DE (equivalently, we reduced the effective sample size in each arm by dividing the original sample size by DE before computing the sampling variance). When ICCs were not reported, we used values from similar studies in the same substantive area, matching outcome type and reporter (self-report, parent, or teacher) as closely as possible (Parker et al., 2025). The impact of these assumptions was examined in sensitivity analyses by varying the ICC over a plausible range.

Meta-analysis

A random-effects model was fitted to the data. The amount of heterogeneity (i.e., τ^2), was estimated using the restricted maximum-likelihood estimator (Viechtbauer, 2005). In addition to the estimate of τ^2, the Q-test for heterogeneity (Cochran, 1954) and the I^2 statistic (Higgins & Thompson, 2002) are reported. In case any amount of heterogeneity is detected, a prediction interval for the true outcomes is also provided (Riley et al., 2011).

For Study II in Saba et al., the descriptive values reported in parentheses in Tables 10 and 11 were not explicitly labelled. In an earlier table in the same paper (Table 7), the column header "M (SD)" makes clear that the parenthetical values represent standard deviations. However, Tables 10 and 11 do not include equivalent column headings, legends, or explanatory notes. As a result, it is unclear whether the values in parentheses were intended to represent standard deviations, standard errors, variances, or another statistical metric.

When these parenthetical values were treated as standard deviations, the resulting SMDs were implausibly large, in some cases reaching values such as Hedges' g = -14.74

Because of this, we did not consider the SD interpretation sufficiently secure for primary analysis. Instead, we conducted sensitivity analyses in which the parenthetical values were alternatively treated as standard errors and as variances.

For each assumption, we calculated post-test between-group Hedges' g using pooled standard deviations and the conventional small-sample correction. These estimates should be interpreted with caution, as the scale of the parenthetical dispersion term could not be verified from the publication. The authors do not explain what the values in parentheses in Tables 10 and 11 represent, nor do they comment on why these values are so small. We also reviewed the Results and Conclusion sections and found no discussion or justification of effect sizes of this magnitude.

On the basis, we recalculated the standardised mean differences under three alternative assumptions, as shown in the table below.

Recalculating the estimates under the assumption that the values were standard errors or variances reduced the magnitude immensely and yielded more moderate effect estimates that were more consistent with the rest of the analysis, and these estimates were therefore used in the main analysis. SEs for each intervention group were converted to standard deviations using:

$$SD = SE \times \sqrt{n}$$

where n is the sample size of the relevant group.

Author description of outcome	SMD if we assume it is a standard deviation	SMD if we assume it is a standard error	SMD if we assume it is a variance
"AN (anger) Dimensions"	-5.70	-1.189	-1.803

"PA (physical aggression) Dimensions"	-2.58	-0.538	-1.616
"VA (Verbal aggression) Dimensions"	-11.14	-2.323	-2.728
"HS (Hostility) Dimensions"	-12.21	-2.547	-3.232
"EC (effort control) dimensions"	6.14	+1.281	+1.229
"surgency dimensions"	-2.62	-0.547	-0.642
"negative affectivity (NA) dimensions"	-14.74	-3.074	-3.297

Publication bias

To assess possible publication bias, effect sizes were first aggregated to a single estimate per study using the aggregate() function in the metafor package, to avoid non-independence arising from multiple effect sizes contributed by the same study. A study-level random-effects model was then fitted using rma.uni() with REML estimation. Publication bias / small-study effects were examined using visual inspection of a funnel plot, Egger's regression test for funnel plot asymmetry, and Begg and Mazumdar's rank-correlation test. Duval and Tweedie's trim-and-fill procedure was used as a sensitivity analysis to estimate the number of potentially missing studies and the corresponding adjusted pooled effect. In addition, a step-function selection model was fitted using selmodel() with two-sided p-value intervals of ≤ .025, .025–.05, .05–.10, and > .10 to examine whether the probability of study inclusion varied according to statistical significance.

Sensitivity analyses

Studentized residuals and Cook's distances are used to examine whether studies may be outliers and/or influential in the context of the model (Viechtbauer & Cheung, 2010). Studies with a studentized residual larger than the $100\times(1-0.05/(2\times k))$th percentile of a standard normal distribution are considered potential outliers (i.e., using a Bonferroni correction with two-sided $\alpha=0.05\alpha=0.05$ for k studies included in the meta-analysis). Studies with a Cook's distance larger than the median plus six times the interquartile range of

the Cook's distances are considered to be influential. The rank correlation test (Begg & Mazumdar, 1994) and the regression test (Sterne & Egger, 2005), using the standard error of the observed outcomes as predictor, are used to check for funnel plot asymmetry. The analysis was carried out using R (version 4.4.2) (R Core Team, 2020) and the metafor package (version 4.8.0) (Viechtbauer, 2010).

Subgroup analysis

Subgroup-specific effects were estimated using robust variance estimation with CR2 adjustment, and statistical significance was assessed using t-tests with Satterthwaite small-sample degrees of freedom. We used coef_test() from the clubSandwich package to compute cluster-robust standard errors (CR2) with Satterthwaite degrees of freedom for hypothesis testing. Differences in effects across outcome domains were examined using a CR2-robust Wald chi-square test.

Results tables presented in the main report show the robust effect size (SMD) of interventions on different subgroup domains, with robust standard errors (SE) and study counts (k = number of effect sizes; n = number of studies). Negative SMDs indicate a reduction/improvement for the intervention group compared to BAU control.

All estimates below are calculated using robust variance estimation (RVE) to account for dependence among effect sizes within studies. p-values correspond to two-tailed tests.

Implementation data

Information on factors that influenced, or were perceived to influence, implementation was extracted from studies where this was reported by study authors.

To capture implementation outcomes the toolkit data extraction made use of Proctor et al's (2011) Implementation Outcomes Framework to capture and categorise the barriers and facilitators to achieving good implementation.

The data extraction for the toolkit is an extension of what is already captured in the EGM. For the EGM, the focus was on whether or not implementation outcomes

were measured. In other words, does a study report on indicators of how well the programme/intervention was implemented or not. For toolkit data extraction we capture why implementation did or did not go well, what influenced implementation? This is typically thought of as barriers and facilitators to implementation. Information on barriers and facilitators will be presented using Proctor et al's (2011) Implementation Outcomes as headings so that the reader can understand the evidence, and gaps in the evidence, on the following implementation outcomes:

- **Acceptability**: Stakeholders' perceptions that the intervention or change is agreeable, palatable, or satisfactory.
    - Example indicators: Children's views on the intervention, participant engagement, satisfaction with content or delivery.
- **Adoption:** The decision or action to employ an intervention or implementation target.
    - Example indicators: Uptake of the intervention by services, schools, or communities.
- **Appropriateness:** The perceived fit or relevance of the intervention to the given context or problem.
    - Example indicators: Adaptations made to improve the intervention's fit with the context, perceived usefulness.
- **Feasibility**: The extent to which the intervention can be successfully implemented in a specific setting.
    - Example indicators: Evidence of practicality or utility, ability to deliver the intervention in the target environment.
- **Fidelity**: The degree to which the intervention was delivered as intended.
    - Example indicators: Training quality, dosage and intensity of the intervention, adherence to the prescribed approach.
- **Reach/Penetration:** The extent to which the intervention has been integrated into a service setting or reached eligible recipients.
    - Example indicators: Ratio of recipients served to the target population, evidence of saturation or integration.
- **Sustainability:** The ability to maintain or institutionalise the intervention over time.

- Example indicators: Evidence of routinisation, integration into policies or practices, durability of implementation efforts.

The information extracted on each implementation outcome was narratively summarised. Where implementation barriers/facilitators or influences on an implementation outcome were not measured and/or reported, this is stated.

Appendix 2. Location Details

	Number of UK Studies	Number (and Location) of International Studies
Contributing to Evidence Quality Rating	2	35 (USA; Turkey; Nigeria; Norway; Netherlands; Denmark; Iceland; Pakistan; Singapore; Spain)
Contributing to Estimated Impact on Violence	2	20 (USA; Turkey; Nigeria; Iceland; Spain; Pakistan; Netherlands)
Contributing to EDIE Data	1	3 (USA; Nigeria)
Contributing to Implementation Data	2	11 (Netherlands; Spain; USA)
Contributing to Cost Data	0	1 (USA)

Appendix 3. Characteristics of included studies for meta-analysis

Authors (Year)	Country	Study Design	Intervention	Population / Place	Delivery Setting	Intervention Provider	Comparison	Outcomes Measured	Quality Level	Findings
Avci (2016)	Turkey	RCT	The Anger Coping Programme aims to help an individual recognise and control anger, particularly when it affects relationships.	Students who struggle with anger management and aggression	School/education setting	Research/study staff	Treatment as usual	Violence Bullying Behavioural difficulties General mental health Regulating and managing emotions	Moderate	Those in the intervention group showed significant improvements, in comparison to the control group, on all outcomes, including psychological symptoms, anger levels and aggression

								Self-esteem		levels, at both post-testing and 6-week follow-up. The only exception was for somatisation, where the improvement was not significant.
Bowman (1982)	USA	RCT	The multimodal treatment approach is designed to modify factors associated with	Male youth offenders	Residential care/social care setting	Research/study staff	Treatment as usual	Behavioural difficulties Regulating and managing emotions	Low	Those in the intervention group showed significantly decreased levels of impulsivity, in comparison to the control group, at post

			impulsive responding.							testing. Additionally, those in the intervention group showed lower-levels of risk-taking post-intervention, compared to the control group, although this difference was not significant.
Burraston (2012)	USA	QED	Class + Cell Phone teaches the Control Model to analyse	Low-to-moderate risk probationers in the	Residential care/social care setting	Research/study staff	Treatment as usual	Crime and offending	Moderate	Results showed that those in the class + cell phone group were

			their behaviour and beliefs and align them accordingly.	juvenile court						significantly less likely to be rearrested, in comparison to the control condition. For those who were rearrested, time to first rearrest was significantly longer for the intervention group, when compared with those in the control group.

Cook (2014)	USA	RCT	Becoming a Man (BAM) + Match Tutoring non-academic supports to teach disadvantaged youth social-cognitive skills and provide academic remediation.	Male youth in 9th or 10th grade with the highest risk scores based on the academic risk index.	School/education setting	Research/study staff	Treatment as usual	Behavioural difficulties School engagement	Moderate	There was a significant improvement in schooling outcomes for those in the intervention group in comparison to the control group. Although there was a reduction in discipline incidents and days suspended for the BAM + Match group, these findings were not significant.

Daunic (2006)	USA	RCT and PE	Tools for Getting Along (TFGA): Teaching Students to Problem Solve is designed to help students develop positive solutions to social problems.	4th and 5th grade students nominated by teachers as being most disruptive or aggressive.	School/education setting	Education/school staff	Treatment as usual	Behavioural difficulties Regulating and managing emotions Violence	High	The intervention was effective at increasing students' foundational procedural knowledge of how to solve social problems; the intervention did not lead to significant changes in how students viewed their own anger or in general teacher ratings of self-control and

										externalizing behaviour; real behavioural prevention may have a "delayed emergence", meaning that positive results might only become apparent in longer-term longitudinal studies; booster lessons did not enhance treatment efficacy; teacher-level factors likely

										influenced how successful students were in learning and applying the skills.
Dogan (2020)	Turkey	RCT	The Roy Adaptation Model (RAM) utilises an holistic approach and focuses on both problematic and healthy behaviours.	Adolescents enrolled in a vocational education centre scored as having high anger scores.	School/education setting	Research/study staff	Treatment as usual	Violence Bullying Behavioural difficulties General mental health Regulating and managing	Moderate	There was a significant difference in anger management levels in the treatment group, suggesting the intervention is clinically effective.

								emotions Self-esteem		
Esposito-Smythers (2017)	USA	RCT and PE	The Alcohol, Self-harm, and HIV Prevention programme (ASH-P) includes planning, assertive communication, cognitive behavioural skills training (e.g., affect management, problem-	13–17-year-olds receiving mental healthcare	Community-based clinical/mental health service	Research/study staff	Treatment as usual	Drug and alcohol use Family relationship and support General mental health Parenting practices	Moderate	Compared to the control group, ASH-P was associated with significantly greater parent-reported openness in communication regarding suicide (medium-large effects), lower odds of deliberate

			solving), and joint parent-teen interaction exercises.							self-harm (large effect), and greater odds of sex refusal at follow-ups. There was also a moderate, though non-significant, reduction in the odds of binge drinking.
Etscheidt (1991)	USA	RCT	Cognitive behavioural training with and without positive reinforcement, adapted from the	Students from a school for "behaviourally disordered youth".	School/education	Education/school setting	Treatment as usual	Violence Behavioural difficulties Regulating and	Moderate	Posttreatment findings showed that both versions of the cognitive-behavioural training

			Anger Control Programme Model, designed to modify aggressive behaviours.					managing emotions		program (training with and without incentives) led to significantly fewer aggressive behaviours and significantly improved teacher-rated self-control compared with the control group. Although the treatment groups did not differ significantly from each

										other on posttreatment measures, the incentive-enhanced condition showed a greater overall degree of improvement from pre- to posttreatment.
Freiden (2006)	USA	RCT	Goal Setting – Anger Management -Manners – Empathy (GAME) targets the behavioural, cognitive	9th–12th grade student with a history of violence	School/education setting	Counsellor	Treatment as usual	Behavioural difficulties School engagement	Low	The intervention (GAME) showed no statistically significant effects on any measured outcome

			and social processes of the participants.							including GPA, school absenteeism, or disciplinary write-ups. Although mean scores suggested small favourable trends for the treatment group (slightly higher GPA gains, fewer days missed, and a modestly larger reduction in write-ups compared

										with controls), none of these changes were significant within groups, between groups, or over time.
Glick (1987)	USA	RCT	Aggression replacement training (ART) is a social skills training programme which targets anti-social behaviours and attempts to replace them with	Boys in a residential facility	Justice/custodial/ community supervision setting	Not stated/unclear	Treatment as usual	Violence Crime and offending Behavioural difficulties Helping others (prosocial	Low	Young people were rated on their post-release community functioning by probation workers. Results showed that those who received ART were rated significantly higher in four

			pro-social behaviours.					behaviours) School engagement Building and maintaining relationships (social-emotional skills) Family relationships and support Opportunities for education,		out of six areas of community functioning (home and family, peer and legal), when compared to those who did not receive ART. There were no significant differences between the groups in two areas (school and work).

								employment, and training Regulating and managing emotions		
Hawkins (1991)	USA	RCT	Behavioural skills training combined with social network development and involvement in prosocial activities.	Adolescents from a juvenile correctional facility.	Justice/custodial/ community supervision setting	Social worker	Treatment as usual	Drug and alcohol use Ability to resolve conflicts Regulating and managing	Moderate	Posttreatment results demonstrated strong training effects on adolescents' skill acquisition and generalisation. There was

								emotions		a significant overall treatment effect, with experimental subjects outperforming controls on total APSI scores and on each individual skill domain: drug and alcohol avoidance, social and problem-solving abilities, and self-control.

Ime (2025)	Turkey	RCT	CBT group counselling programme which aims to reduce bullying and enhance adolescents ' empathy skills.	Adolescents exhibiting bullying behaviour	School/education setting	Mental health/therapeutic professional	Treatment as usual	Violence Bullying Building and maintaining relationship (social-emotional skills)	Moderate	Those in the CBT group counselling programme showed significantly lower bullying perpetration scores and significantly higher empathy scores at post-testing in comparison to the control group. These results were maintained at a 2-month follow-up.

Johnsen (2024)	Denmark	RCT	Back2School aims to increase school attendance by involving youth, parents and school personnel.	6–16-year-olds with more than 10% school absence in previous three months	Community-based clinical/mental health setting	Mental health/therapeutic professional	Treatment as usual	Behavioural difficulties School engagement Parenting practices Self-esteem	Moderate	Although school attendance increased in both groups, there was not a significant difference between the treatment and control. The treatment group improved by a greater margin on the SDQ and emotional symptoms scales. On the problems with peers

										scale, the treatment group improved, while the control group worsened.
Karatas (2009)	Turkey	RCT	CBT to reduce aggression.	9th grade students with high-level aggression.	School/education setting	Education/school staff	Treatment as usual	Violence Behavioural difficulties	Low	There was a significant difference between the post-test total aggression, physical aggression, hostility and indirect aggression scores in the control and

										treatment groups.
Kettlewell (1983)	USA	RCT	CBT using behavioural rehearsal and self-instruction training to reduce aggressive behaviour.	Children who had one recorded incidence of verbal or physical aggression at a summer day camp each day before baseline.	Residential care/social care setting	Not stated	Treatment as usual	Violence Behavioural difficulties	Moderate	When combining all outcome measures, the treatment group improved significantly more than the control. Specific outcomes with significant improvements included counsellor ratings of physical and

										verbal aggression.
Kumuyi (2022)	Nigeria	RCT	CBT using cognitive restructuring.	10–18-year-olds attending secondary school with a severe level of conduct disorder	School/education setting	Mental health/therapeutic professional	Treatment as usual	Violence Behavioural difficulties	Moderate	There was a significant difference in the decrease of conduct disorder in favour of the treatment group when compared with the control.
LeSure-Lester (2002)	USA	RCT	A cognitive behavioural skilled group designed to teach skills to use in	African American male 12-16-year-olds living in a group home	Residential care/social care setting	Mental health/therapeutic professional	Treatment as usual (indirect therapy)	Violence Behavioural difficulties	High	Youth in the treatment group demonstrated greater behaviour compliance and less

			aggressive situations.							aggression towards staff at endline when compared with the control group.
Lochman (2017)	USA	RCT and PE	The Coping Power — Internet Enhanced (CP-IE) programme teaches children social problem-solving, goal-setting and emotional regulation skills, and	4th grade children exhibiting aggressive behaviour (with participation in 5th grade)	School/education setting	Mental health/therapeutic professional	Treatment as usual	Violence Behavioural difficulties	High	Those in the CP-IE intervention group displayed significantly lower increases in conduct behaviour problems over the 5th grade year compared to control children. At

			teaches parents parenting skills.							post-intervention, 17% of children in the intervention group were within the clinical range for conduct problems, compared to 35% of children within the control group. However, there were no significant differences between the two groups on teacher-rated

										aggression problems.
Logsdon (2004)	USA	RCT and PE	Cognitive behavioural group designed to reduce aggression in interpersonal relationships.	Boys aged 8-11 referred by teachers due to aggressive behaviour problems	School/education setting	Mental health/therapeutic professional	Treatment as usual	Behavioural difficulties Ability to resolve conflicts	Low	There were no significant differences found between the intervention and control groups on any outcome measures, at post-testing. The author does note that due to small sample sizes they are unable to draw any conclusions regarding the

										intervention's efficacy.
McLaughlin (2009)	USA	RCT	A cognitive behavioural bullying prevention intervention with an added media component to decrease bullying and victimisation, and increase empathy and bystander response.	6th grade students who are potential bullies, victims, or bystanders	School/education setting	Research/study staff	Treatment as usual	Violence Bullying Helping others (prosocial behaviours) Victim of crime Building and maintaining relationships (social-emotional skills)	Moderate	There was improvement in the victimisation, bullying and bystander scores in the intervention group, but this was not statistically significant.

Moynahan (2005)	Norway	RCT	Aggression Replacement Training (ART) with enhancements designed to strengthen social competence and decrease aggressive behaviour.	7-12 and 14-20-year-olds, some with ASD and ADHD	Research/study staff	Education/school staff	Treatment as usual	Behavioural difficulties Building and maintaining relationships (social-emotional skills)	Low	For the adolescent ART group, there were no significant changes from pre-test to post-test across domains. For the children ART group, there was a statistically significant improvement in social skills and a significant decrease in problem behaviours.

Njardvik (2022)	Iceland	RCT	Tuning Your Temper aims to reduce disruptive behaviour problems, particularly through becoming aware and learning ways to control their emotional reactions.	Children referred to psychological services due to behavioural problems	School/education setting	Mental health/therapeutic professional	Treatment as usual	Violence Behavioural difficulties	Moderate	There was a statistically significant decrease in teacher rated conduct problems and ODD symptoms, which was maintained at the 6-month follow-up. However, parent rated findings were only significant at the follow-up. There was no control group measure at the follow-up.

Nwolisa (2023)	Nigeria	RCT	CBT (cognitive restructuring) and social learning therapy (modelling) to better think about problems and observe real life models.	Adolescents	Justice/custodial/ community supervision setting	Research/study staff	Treatment as usual	Violence Behavioural difficulties General mental health Self-esteem	Moderate	Both interventions were effective in reducing aggression and anxiety when compared to the control group, although depression and low self-esteem were reduced more in the social learning therapy group.

Oparaduru (2017)	Nigeria	QED	Cognitive Restructuring Therapy is a training package on cognitive restructuring and self-control, designed to tackle maladaptive behaviour.	Senior Secondary School Two (16-17-year-olds) scoring highly on a psychopathy checklist.	School/education setting	Research/ study staff	Treatment as usual	Behavioural difficulties Self-esteem	Moderate	There was a statistically significant improvement for the treatment groups in maladaptive behaviour and self-esteem. However, the self-control therapy group performed better than the cognitive restructuring group.

Orgilés (2023)	Spain	RCT and PE	Super Skills for Life (SSL) was adapted for online implementation. It is a transdiagnostic programme based on CBT, designed to tackle risk factors for children's emotional problems.	8-12-year-olds displaying emotional symptomatology.	Computerised/digital	Mental health/therapeutic professional	Treatment as usual	Violence Behavioural difficulties Helping others (prosocial behaviours) Meaningful relationships General mental health Regulating and managi	Moderate	The intervention had a statistically significant effect in seven of the ten reported outcomes, including internalising and externalising problems, in both the children's and parents' reports. The effect sizes were considered to be large.

								ng emotions		
Poulin (2001)	USA	QED and PE	Adolescent Transitions Programme (ATP) is a curriculum to tackle adolescents' regulation of their prosocial and disruptive behaviour in settings with parents and peers.	11-14-year-olds considered to be 'high-risk' based on 10 dimensions of child risk	Community-based clinical/mental health service	Research/study staff	Treatment as usual	Behavioural difficulties Drug and alcohol use	Moderate	Results showed that, in comparison to the control groups, there was a significant increase at post-testing in frequency of smoking for young people in the peer intervention groups, with this effect still present 3 years after

										the intervention. Similar results were found for delinquency, there were no significant effects of the intervention at post testing, however there was a significant difference, favouring the control group, at 1-year follow up. This was also the case at 2 and 3-year

										follow-ups, although the difference was not significant.
Pullen (1996)	USA	RCT and PE	Juvenile Intensive Supervision Probation (JISP) - Reasoning and Rehabilitation (R&R) cognitive skills development programme is a series of modules which aims to teach	Male adolescents placed on Juvenile Intensive Supervision Probation.	Justice/custodial/ community supervision setting	Research/study staff	Treatment as usual	Crime and offending	Very low	Participants in the treatment group were twice as likely as those in the control group to commit a technical violation during the programme, but participants in the control group were three times

			cognitive skills to offenders.							more likely after the programme ended. The author reports that the differences in recidivism rates post treatment are not statistically significant, suggesting that participation did not affect crime and offending.

Saba (2023)	Pakistan	RCT	CBACT aims to reduce aggression and behaviour difficulties among primary school students by teaching through animations.	7-9-year-olds with a high level of anger, aggression, and behavioural issues	School/education setting	Research/study staff	Treatment as usual	Violence Behavioural difficulties General Mental Health Regulating and managing emotions	Low	The treatment group experienced better outcomes on the aggression and behaviour scales.
Schwartz-Mette (2024)	USA	RCT and PE	The FRIENDS Resilience programmes aim to reduce physiological, cognitive,	Kindergarten, 2nd, 5th and 7th grade students	School/education setting	Mental health/therapeutic professional	Treatment as usual	Behavioural difficulties School engagement	Moderate	There was not a significant difference in outcomes between the treatment and control

			and behavioural maladaptation.					Building and maintaining relationships (social-emotional skills)		groups in social skills, problem behaviours, academic competence.
Squires (2012)	England	RCT	Small group CBT for adolescents with externalising behavioural difficulties, focusing on self-regulation and problem-	Adolescents with specific behavioural needs within a mainstream school context, excluding those with a formal	School/education setting	Mental health/therapeutic professional	Treatment as usual	Violence Behavioural difficulties Regulating and managing emotions	High	Pupils in the CBT group reported a significant improvement in their self-perceptions of inattention and hyperactivity, yielding a large effect size compared to

			solving skills.	clinical diagnosis.						the control group, who viewed their own behaviour as worsening. In contrast, teacher ratings indicated that the behaviour of all pupils in both groups improved over the study period, resulting in no significant differential effect between the treatment

										and control conditions.
Sukhodolsky (2000)	USA	RCT and PE	Cognitive-behavioural treatment to better control anger and reduce the frequency and intensity of this anger.	4th and 5th grade boys nominated by teachers as having "anger-related problems"	School/education setting	Mental health/therapeutic professional	Treatment as usual	Violence Behavioural difficulties	Moderate	Teachers viewed the treatment group to be less aggressive and disruptive when compared to the control. However, self-reported data suggests the intensity of angry feelings was not reduced.

Te Brinke (2021)	Netherlands	RCT and PE	Think Cool; Act Cool targets emotion regulation for those with externalising problems through a cognitive module ('Think Cool') and behavioural module ('Act Cool').	Adolescents with a teacher-reported subclinical or clinical level of externalising problems and an IQ over 80	School/education setting	Mental health/therapeutic professional	Treatment as usual	Violence Behavioural difficulties Regulating and managing emotions	Moderate	There was a significant overall effect of the intervention on adaptive emotion regulation strategies. However, there were no significant effects on emotion regulation difficulties or maladaptive emotion regulation strategies. There was also a effect on self-

										reported externalising problems in favour of the control group. However, there was no significant effect on the outcomes reported by parents and teachers.
Wagner (2014)	USA	RCT	Guided Self-Change (GSC) is designed to reduce substance abuse and aggressive behaviour.	14-18-year-olds with six occasions of alcohol or drug use in the previous 90 days	School/education setting	Mental health/therapeutic professional	Treatment as usual	Violence Behavioural difficulties Drug and alcohol use	Moderate	The treatment group had significantly fewer alcohol and drug use days, and reports of aggressive behaviour

										than the control group. This effect was maintained at the three-month follow-up but was no longer present by the six-month follow up.
Watson (2009)	UK	QED	Primary Anger Management, key stage 2 version: Learning how to deal with out angry feelings is designed to	7-11-year-olds who found managing their anger difficult	School/education setting	Mental health/therapeutic professional	Treatment as usual	Behavioural difficulties Helping others (prosocial behaviours)	Low	There was no significant difference in SDQ scores and subscale scores between baseline and endline for either the treatment or

			improve anger management in primary school children.					School engagement Regulating and managing emotions		control group. Furthermore, there was not a significant difference between the two groups. The author deemed the intervention to not be successful.
Yeo (2011)	Singapore	RCT	A psychoeducational CBT programme designed to manage behavioural difficulties, anger issues, and difficulties	8-12-year-olds exhibiting behavioural difficulties and being disruptive in class.	School/education setting	Mental health/therapeutic professional	Treatment as usual	Behavioural difficulties Building and maintaining relationships	High	The treatment groups made marked, statistically significant improvements compared to the control groups across all six

			at school by increasing self-awareness, self-esteem and self-control.					(social-emotional skills) Regulating and managing emotions Self-esteem		domains examined: self-control, classroom behaviour, school behaviour, home behaviour, social skills and interpersonal relationships, and self-esteem. However, self-esteem was the only outcome maintained at the one-month follow-up.


YOUTH
ENDOWMENT
FUND


NATIONAL
CHILDREN'S
BUREAU

Appendix 4. Characteristics of included studies for implementation

Authors (Year)	Country	Study Design	Intervention (Name and brief description)	Quality Level	Summary of Findings
Augustyniak (2009)	USA	RCT	**The Prepare Curriculum:** an integrated cognitive-behavioural/ interpersonal skills approach in a weekly psychoeducational group format	Low	**Feasibility:** Group facilitators delivered the intervention to 10 groups in 10 school districts using a 10-week protocol that was specifically selected because it was conducive to a school setting. **Fidelity:** Facilitators received training on The Prepare Curriculum. Treatment integrity specific to the treatment timeline and the application of the curriculum was monitored via periodic follow-up by the authors and was deemed to be "excellent".
Cole et al. (2013)	England	RCT and Process evaluation	**Learning How to Deal with our Angry Feelings:** a short-term cognitive-behavioural programme for groups of primary-age	Moderate	**Fidelity:** Fidelity of implementation was high, with co-facilitators rating sessions an average of 4.76 out of 5 across all 13 groups, and no

			children with anger-related difficulties		significant differences observed between cohorts or facilitators.
Cook et al. (2014)	USA	RCT	**Becoming a Man (BAM):** social-cognitive skills training based on the principles of CBT	Moderate	**Fidelity:** The intervention was manualised and dosage tracked through provider records. Authors reported on potential treatment crossover and spillover effects between experimental groups in school setting. **Cost:** The cost per participant of the integrated non-academic and academic intervention was estimated at approximately $4,400. The authors found that the intervention yielded larger gains in outcomes per dollar spent compared with many prominent alternatives for disadvantaged young people.
Daunic et al. (2006)	USA	Process evaluation	**Tools for Getting Along (TFGA):** an intervention designed to help 4th and 5th grade students to develop positive solutions to social problems, particularly in anger-provoking situations	High	**Acceptability:** Teachers reported generally positive perceptions of the TFGA program, indicating that it was appealing, useful for students, and something many would be willing to implement again in the future. **Appropriateness:** The curriculum was viewed as suitable for classroom settings, with age-

					appropriate content and a structured 30-minute lesson format that could be integrated into regular teaching schedules. **Fidelity:** Most teachers reported delivering the full lesson content and activities as intended, and classroom observations largely confirmed adherence to the curriculum structure and key components
Esposito-Smythers et al. (2017)	USA	RCT	**Alcohol, Self-Harm and HIV Prevention Programme (ASH-P):** cognitive behavioural family-based alcohol, self-harm and HIV prevention programme for young people in mental health treatment	Moderate	**Feasibility:** Practical constraints with study recruitment and retention rates were highlighted. The length and time commitment of the 12-hour workshop (held over two weekends) was identified as a barrier. **Fidelity:** The intervention was delivered according to a protocol. The study reported 100% adherence to the core components of the protocol. The therapists' competence of delivery was rated as having been done well or very well in over 90% of the delivery.
Lochman et al. (2017)	USA	Process evaluation	**Coping Power - Internet Enhanced (CP-IE):** an	High	**Acceptability:** Children and parents reported high satisfaction with the CP-IE programme,

			indicated prevention program for preadolescents who are at-risk because their problem behaviours can be precursors for later severe conduct problems, substance abuse, and juvenile delinquency		particularly valuing the face-to-face sessions, while children also found the website engaging and helpful. **Appropriateness:** The hybrid structure and reduced number of in-person sessions fit well within school settings, allowing the programme to integrate with school schedules while maintaining the core Coping Power content. **Feasibility:** The programme was feasible to implement in schools, with successful integration of sessions into the school day, use of the online platform for ongoing activities, and manageable training and monitoring procedures for staff. **Fidelity:** Intervention delivery demonstrated high fidelity, supported by structured manuals, trained clinicians, weekly supervision, and self-reported completion of session objectives.
Logsdon (2003)	USA	RCT	**Cognitive Behavioural Group:** CBT to reduce	Low	**Adoption:** Only two school systems agreed to participate. There was resistance from school officials and institutional barriers.

			aggressive behaviour and delinquency		**Appropriateness:** Adaptations were made to make the intervention suitable for aggressive elementary-aged children, including aligning with development levels and engagement needs.
Orgilés et al (2023)	Spain	RCT	**Super Skills for Life (SSL):** self-applied computerised programme designed to target children with emotional problems	Moderate	**Appropriateness:** The programme was adapted to a digital format to reach other children. Objectives were retained but activities were computerised.
Pullen (1996)	USA	RCT and PE	**Reasoning and Rehabilitation:** a cognitive skills development programme as part of the Juvenile Intensive Supervision Probation (JISP)	Very low	**Appropriateness:** Recommendations were made for shortening the duration of the programme to better suit the attention spans of the participants. **Fidelity:** The sessions were not always delivered as intended, with preparation for sessions being of particular concern. **Reach/penetration:** The officers and trainers remarked that it was easier to recruit adults than young people to the programme.

Schwartz-Mette et al. (2024)	USA	RCT	**FRIENDS Resilience Programmes:** cognitive-behavioural skills intervention tailored to different age stages, both a universal and selective prevention tool	Moderate	**Fidelity:** Sessions were delivered by trained and supervised FRIENDS facilitators. Authors observed 30% of sessions and rated whether objectives were met. 99.2% of session objectives were met, 98.6% of scheduled activities completed and 100% of at-home activities assigned.
Te Brinke et al. (2021)	Netherlands	RCT	**"Think Cool; Act Cool":** cognitive and behavioural modules to tackle externalising problems and emotion regulation	Moderate	**Fidelity:** Reviewing audiotaped sessions showed high treatment fidelity.
Wagner et al. (2014)	USA	RCT	**Guided Self-Change (GSC):** a school-based brief motivational interventions and CBT intervention	Moderate	**Fidelity:** Treatment fidelity was believed to be high after reviewing session recordings.
Watson (2009)	UK	QED	**Cognitive Behavioural Therapy:** CBT to target anger management	Low	**Acceptability:** Participants noted improvements in their anger management and were positive about the group.

					Feasibility: There were logistical barriers and disruptive behaviour which made the intervention more difficult to implement. **Fidelity:** The facilitators were rated as having high treatment fidelity when delivering the programme.

Appendix 5. Availability of evidence according to each of Proctor et al.'s (2011) implementation outcomes

Authors (Year)	Acceptability	Adoption	Appropriate-ness	Feasibility	Fidelity	Reach/ penetration	Sustainability	Cost
Augustyniak (2009)	No	No	No	Yes	Yes	No	No	No
Cole et al. (2013)	No	No	No	No	Yes	No	No	No
Cook et al. (2014)	No	No	No	No	Yes	No	No	Yes
Daunic et al. (2006)	Yes	No	Yes	No	Yes	No	No	No
Esposito-Smythers et al. (2017)	No	No	No	Yes	Yes	No	No	No
Lochman et al. (2017)	Yes	No	Yes	Yes	Yes	No	No	No
Logsdon (2003)	No	Yes	Yes	No	No	No	No	No
Orgilés et al (2023)	No	No	Yes	No	No	No	No	No
Pullen (1996)	No	No	Yes	No	Yes	Yes	No	No
Schwartz-Mette et al. (2024)	No	No	No	No	Yes	No	No	No

Te Brinke et al. (2021)	No	No	No	No	Yes	No	No	No
Wagner et al. (2014)	No	No	No	No	Yes	No	No	No
Watson (2009)	Yes	No	No	Yes	Yes	No	No	No

Appendix 6. Moderator categories

Setting-level moderators

Setting

Category	Description
School / education setting	Intervention delivered in or recruitment drawn from schools or general education settings, including primary, secondary, specialist or vocational schools.
Community-based clinical / mental health setting	Intervention delivered in or recruitment drawn from non-residential health, mental health, counselling, or outpatient community services.
Residential care / social care setting	Intervention delivered in or recruitment drawn from group homes, residential care, child welfare placements, or other non-custodial care settings.
Justice / custodial / community supervision setting	Intervention delivered in or recruitment drawn from secure or custodial youth justice settings, such as correctional centres, detention facilities, or residential juvenile justice placements AND Intervention delivered in or recruitment drawn from youth justice services in the community, such as probation, court supervision, or intensive supervision programmes.

Computerised / digital delivery format	Intervention delivered through a computerised, digital, or technology-based format rather than solely face-to-face methods.

Intervention-level moderators

Inclusion criteria

Category	Description
Justice involved / offending-related CYP	Children and young people included because they are involved in the youth justice system or have an offending history, including custody, probation, court supervision, correctional placement, or documented offending/violence records.
Externalising / behavioural dysregulation	Children and young people included because of aggression, anger, bullying, disruptive behaviour, impulsivity, conduct problems, acting-out behaviour, oppositional behaviour, or related externalising difficulties.
Mental health / emotional difficulties	Children and young people included because of emotional symptoms, psychological distress, mental health need, psychiatric difficulty, or receipt of mental health support, where the primary inclusion criterion is not specifically externalising behaviour.
School-risk / disengagement profile	Children and young people included because of school attendance problems, absenteeism, exclusions, academic failure, being over-age for grade, disengagement from school, or other education-based risk indicators.

Social care / adversity-exposed CYP	Children and young people included because of care experience, child welfare involvement, abuse history, group home or residential care placement, protective services involvement, or other significant adversity exposure.
Universal / whole-population sample	Children and young people included because they were part of a whole class, year group, school, or other universal population, without targeted selection based on individual risk or need.

Family Involvement

Category	Description
Integrated Joint Family Sessions (High Involvement)	The family system itself is treated as a core component of the intervention. Parents and children attend therapy sessions together to actively collaborate, role-play, and practice skills under the guidance of a therapist. The focus is on interactive problem-solving, creating shared risk-prevention plans, and improving parent-child communication dynamics in real-time
Parallel Parent Management Training (Moderate Involvement)	Parents actively participate in their own structured, multi-week training sessions that run concurrently with the child's CBT groups. The focus is on altering the child's environment by teaching parents specific family management skills, such as positive reinforcement, consistent limit-setting, parental monitoring, and alternative discipline strategies. This enables parents to actively reinforce the "positive opposites" of their child's problem behaviours.

Informational Support and Homework Facilitation (Low Involvement)	Parents in this category act as supportive observers who help reinforce the CBT concepts outside of the therapy room. Programs provide parents with newsletters, summary sheets, or access to online portals that explain the skills their child is learning that week. Parents are encouraged to prompt their child to use these skills or help them complete assigned "homework" tasks, but the parents do not receive formal clinical training themselves.
No family component	Parents have no active therapeutic role in the intervention. Their involvement is either not mentioned at all or strictly limited to the administrative/logistical requirements of the study or program, such as providing legal consent for their child to participate and completing baseline or post-intervention behavioural assessments (e.g., the Child Behavior Checklist or Strengths and Difficulties Questionnaire). The CBT training is entirely child-focused and delivered directly to the youth

Intensity

Category	Description
Low Intensity / Brief Interventions	These interventions are generally short in duration, ranging from a few hours total to around 5-8 sessions. They are highly targeted, often focusing on a single skill set (like cognitive restructuring or basic anger management) and require minimal time and resources to implement.

Moderate Intensity / Standard Interventions	These interventions represent a standard clinical or school-based approach. They typically span 10 to 20 sessions (roughly 12 to 15 hours of contact time) delivered over a few months. They occasionally integrate a secondary component, such as booster sessions or brief parent involvement.
High Intensity / Comprehensive & Multi-Systemic Interventions	These interventions require a massive time commitment, spanning dozens of sessions, an entire academic year, or multiple years. They often utilise a multi-tiered approach (universal classroom + targeted groups + family therapy) or combine intense behavioural therapy with daily academic remediation.

Individual or group, delivery method

Category	Description
Individual, face to face	Intervention delivered one-to-one in person between the practitioner and the child or young person, without a peer group component.
Small groups (2–5), face to face	Intervention delivered in person to a small group of two to five participants at a time.
Medium groups (6–14), face to face	Intervention delivered in person to a mid-sized group of six to fourteen participants at a time.
Large groups (15+), face to face	Intervention delivered in person to a large group, class, or workshop-style format involving fifteen or more participants.

Groups (size unclear), face to face	Intervention delivered in person in a group format, but the number of participants per group is not reported or cannot be determined.
Individuals and small groups (5 per group), face to face	Intervention delivered in person using a mixed format that combines one-to-one sessions with small-group sessions of up to around five participants.
Hybrid interventions (face-to-face + technology/media)	Intervention delivered through a combination of in-person contact and a technology- or media-based component, such as phone, computer, internet, app, video, or other digital/self-directed element.
Individual, and face to face delivery	In this format, the intervention is delivered directly to the individual youth (or the youth and their specific family unit) without the presence of peers.

Population-level moderators

Ethnicity

Category	Description
100% Black or Global Majority Sample	Entire sample identified as Black and Global Majority (i.e., Black, Hispanic/Latinx, Asian, Indigenous, etc.)
Majority white Sample	>85% of sample identified as white

Some Diversity	Between 15–49% of participants identified as Black and Global Majority (i.e., Black, Hispanic/Latinx, Asian, Indigenous, etc.)
Diverse / Balanced	50% or more Black and Global Majority youth in the sample
Not stated / unclear	Study does not specify the ethnicity of participants

Sex/Gender

Category	Description
All male	Sample is entirely male
All female	Sample is entirely female
Majority male	Sample included >70% male
Majority female	Sample included >70% female
Mixed sex	40–70% male and 30–60% female
Not stated / unclear	Study did not report the gender breakdown of participants

Outcome moderators

Outcome type

Category	Description
Ability to resolve conflicts	Skills to de-escalate conflicts and arguments.
Behavioural difficulties	Disruptive or aggressive behaviours due to distress or needs, excluding hyperactivity and neuro-diverse conditions.
Building and maintaining relationships	Social-emotional skills for listening, cooperating, and understanding emotions.
Crime and offending	Criminal behaviours including non-violent acts (e.g., shoplifting), sexually violent crimes, and violent acts (e.g., assault, robbery).
Drug and alcohol use	Problematic use of substances causing harm to self or others, impacting health, school, or social functioning.
Family relationships and support	Positive and supportive connections with family members.
General mental health	Encompasses psychological mental health, which specifically relates to an individual's cognitive and emotional functioning, including their ability to manage thoughts, behaviours, and cope with life's challenges. Includes mental health conditions like anxiety, depression, PTSD

Helping others (prosocial behaviour)	Engaging in positive, prosocial behaviours such as helping, comforting, or sharing.
Parenting practices	Parenting practices, often learnt from a parent/carer's own experience of being a child, that are not appropriate for a certain situation. It includes harsh or inappropriate discipline, controlling behaviour, inconsistent parenting, or low parental warmth.
Regulating and managing emotions	Also called 'emotion regulation'. Having the skills and techniques to manage feelings and reactions to situations and events, reducing the intensity, duration, and impact of such feelings.
Self-esteem	Viewing yourself positively, including confidence in your abilities, appearance, and self-worth.