Review

Cross-cultural adaptability of parenting interventions designed for childhood behavior problems: A meta-analysis

Laura Maciel [a,*], Aitana Gomis-Pomares [b], Crispin Day [c], Miguel Basto-Pereira [a]

[a] *William James Center for Research, Ispa-Instituto Universitário, Rua Jardim do Tabaco 34, 1149-041 Lisbon, Portugal*
[b] *Developmental Psychology Department, Universitat Jaume I, Avinguda de Vicent Sos Baynat, s/n, 12006 Castelló de la Plana, Castelló, Spain*
[c] *Department of Psychology, Child & Adolescent Mental Health Service Research Unit, King's College London, Institute of Psychiatry, Psychology and Neuroscience, and Centre for Parent and Child Support, South London and Maudsley NHS Foundation Trust, Michael Rutter Centre, 16 De Crespigny Park, London SE5 8AF, United Kingdom*

A B S T R A C T

The dissemination of parenting interventions is one of the advised approaches to globally counteract childhood behavior problems, delinquency, and future criminal careers. Many of these interventions are developed in Anglosphere countries and transported to other contexts with distinct cultural backgrounds. However, there are no meta-analyses evaluating the overall effectiveness of these Anglosphere parenting programs in non-Anglosphere settings. This meta-analysis aimed to examine the effectiveness of parenting interventions developed in Anglosphere countries when transported to non-Anglosphere countries, as well as compare effectiveness levels between Anglosphere and non-Anglosphere trials; and analyze the impact of research and contextual factors in the dissemination of these interventions. Parenting interventions were included if they were: created in an Anglosphere setting; tested in non-Anglosphere countries; focused on reducing childhood behavioral problems; designed for children ranging from two to 12 years old; and tested in an experimental randomized trial. A random-effects model was selected for our meta-analysis. Standardized mean differences, confidence intervals and prediction intervals were also computed. Twenty studies were included, and results suggest that parenting interventions designed for childhood behavior problems can be transported to non-Anglosphere countries and potentially maintain effectiveness. This study is a relevant contribution to the evidence of cross-cultural transportability of parenting interventions.

1. Introduction

Parenting interventions are known to be one of the most effective strategies when dealing with childhood behavior problems (e.g., Florean, Dobrean, Pasarelu, Georgescu, & Milea, 2020; Jeong, Franchett, de Oliveira, Rehmani, & Yousafzai, 2021). They are renowned for their immediate and long-term outcome benefits, which encompasses improvements in parent well-being and family functioning (e.g., Barlow & Coren, 2017; Smith et al., 2020), reductions in future antisocial behavior (e.g., Basto-Pereira & Farrington, 2022; Farrington, Gaffney, Lösel, & Ttofi, 2017), and the promotion of adequate developmental pathways for children (e.g., Pedersen et al., 2019).

Indeed, the United Nations Guidelines for the Prevention of Juvenile Delinquency (General Assembly Resolution 45/112, 1990) underline the

importance of family-based interventions in order to counteract delinquency and criminal behavior. They state that evidence-based programs can "provide families with the opportunity to learn about parental roles and obligations as regards child development and childcare, promoting positive parent-child relationships, sensitizing parents to the problems of children (…) and encouraging their involvement in family and community-based activities" (General Assembly Resolution 45/112, 1990, p. 202), thus preventing antisocial behavior.

Therefore, there has been a rising surge of policy recommendations for the implementation of parenting programs around the globe. According to several governments and international organizations, such as the World Health Organization (WHO, 2010) and the United Nations Office on Drugs and Crime (UNOFC, 2009), the supported method is to transport interventions that are evidence-based and have already proven

* Corresponding author.
E-mail addresses: lauramaciel09@outlook.com (L. Maciel), apomares@uji.es (A. Gomis-Pomares), crispin.1.day@kcl.ac.uk (C. Day), mbpereira@ispa.pt (M. Basto-Pereira).

their efficacy to other countries.

The dissemination of parenting interventions to other cultural contexts may have various advantages, especially if we take into account that developing an intervention can be costly, either in material, human or time resources (Leijten, Melendez-Torres, Knerr, & Gardner, 2016). In addition, when an intervention has proven its effectiveness in a given context, it is feasible to believe that it can be effective in another environment (Gardner, Montgomery, & Knerr, 2016). This seems to be due to the fact that basic principles of parenting interventions may be universal (Lunkenheimer, Olson, Hollenstein, Sameroff, & Winter, 2011) like, for example, their theoretical background. Most parenting interventions designed for reducing behavior problems in children were created based on the operant learning theory and social learning theory, which are theories that have established their validity across different cultures (Leijten et al., 2016), and therefore are believed to be less affected by cultural factors. This hypothesis seems to be supported by the multilevel meta-regression study from Leijten et al. (2016). The goal of this study was to compare effectiveness of transported and home-grown interventions from different regions in reducing childhood behavior problems. Based on their results, the authors concluded that parenting interventions that were created with the same theoretical principals and components lead to similar outcomes, independently of being transported or homegrown.

Notwithstanding the above, many specialists underline that culture and context play a major role in the development and dissemination of parenting programs (e.g., Martinez et al., 2020). They argue that, although core components of parenting interventions are indeed universal, specific parenting techniques may have distinct impacts on childhood behavior problems, accordingly to culture. Physical punishment, for example, seems to have a differential impact on disruptive and delinquent behavior depending on ethnicity (Sahithya, Manohari, & Vijaya, 2019; Simons et al., 2004). The same phenomenon seems to happen to descriptive praise. Although descriptive praise is a universally accepted parenting strategy to prevent behavioral problems, direct and indirect praise seem to present distinct results in western and eastern populations (Fong, 2010; Hill & Tyson, 2008). Morelli et al. (2018) also make a relevant argument when discussing positive parenting practices in low to middle-income countries. They state that positive parenting practices (very frequently present in all parenting interventions) are based in attachment theories, which is a theory based in the way of living and the quality of life of Western communities, which has many disparities from middle to low-income countries (Morelli et al., 2018). Furthermore, what is perceived as appropriate and inappropriate parenting practices can vary immensely depending on cultural backgrounds (Gonzales, 2017; Heim & Kohrt, 2019). In this context, some parenting interventions may end up encouraging parents to change their behaviors and perspectives, resulting in a negative impact to the child, the family and their participation in their social groups.

Therefore, a debate has risen in the empiric literature regarding the need for cultural adaptation of mental health interventions. Some authors support the argument that modifications to the original intervention, which is evidence-based, can impact the delivery of core components responsible for mechanisms of change (Elliot & Mihalic, 2004). In this sense, only by rigorous fidelity and guaranteeing adherence to the original intervention can we assure efficacy and effectiveness (Elliot & Mihalic, 2004). In opposition to this argument, some researchers (e.g., Castro, Barrera, & Holleran Steiker, 2010; Meija, Leijten, Lachman, & Parra-Cardona, 2017) state that an intervention needs to be in agreeance with cultural norms, values and experiences of target populations in order to be effective, highlighting thus the term of cultural relevance. Hence, only when cultural relevance is accounted for, successful implementation and outcome changes can be ensured (Bernal, Bonilla, & Bellido, 1995; Castro et al., 2010), and so these need to be accounted for with particularly attention when transporting mental health interventions (Meija et al., 2017).

This debate becomes even more important in this area of expertise

when we take into account that, similarly to >70% of psychological research (Rad, Martingano, & Ginges, 2018), many parenting interventions originate from Western, educated, industrialized, rich, and democratic (WEIRD) nations. More specifically, they are developed in Anglosphere countries (i.e., English-speaking, developed nations that share common cultural and historical ties to the United Kingdon), which is the case of well-known interventions like Incredible Years (Webster-Stratton, 2001), Triple P (Sanders, 1999) and Parent-Child Interaction Therapy (PCIT; Eyberg, Boggs, & Algina, 1995). Subsequently, these are also the interventions targeted to transportation to other cultures. This might be due to the fact that WEIRD contexts have more socioeconomic opportunities to create, test and disseminate parenting interventions (Cheon, Melani, & Hong, 2020). Moreover, some argue that WEIRD samples and sociocultural backgrounds are more representative of human behavior (Maryanski, 2010). However, several authors (e.g., Henrich, 2020) have counterargued this hypothesis and defended the idea that WEIRD settings are the least representative, since they seem to be outliers in many fields of social sciences. Thus, this raises concerns regarding the adaptability of programs created in WEIRD contexts to countries that have distinct cultural values and socioeconomic backgrounds (Morelli et al., 2018).

Considering the aforementioned, several cultural adaptation frameworks have been developed in order to ensure that interventions are adapted in order to be in accordance with the cultural values and daily experiences of the target population (Gonzales, 2017). According to Baumann et al. (2015), there are two categories of cultural adaptation models, namely those that focus on modifying the intervention content, and those that focus on the process of adaptation. Regarding the first set of frameworks, the Ecological Validity Model (EVM; Bernal et al., 1995) and the Cultural Sensitivity Model (Resnicow, Soler, Braithwaite, Ahluwalia, & Butler, 2000) are the most discussed in the empirical literature (e.g., Heim & Kohrt, 2019). They are employed in order to adapt the intervention manual and its delivery methods, materials, or even cultural or historical factors that may influence the target population (Meija et al., 2017). In contrast, the second set of models are centered on the decisions related to the process of adaptation (e.g., when and how to adapt, which participants are to be involved) and are generally based on formal assessments of the adapted intervention (Cooper et al., 2019). They usually focus on setting out a number of guidelines and recommendations a priori, in order to help guide the process of adaptation (Baumann et al., 2015). One example of this framework is the Cultural Adaptation Process (CAP; Domenech-Rodríguez & Wieling, 2004). This model, with the intent of complementing the EVM, it highlights the importance of collaborative relationships between program developers, researchers, policy makers, practitioners, and participants. Furthermore, it focuses on testing the adapted interventions and gather information to better tune the adapted intervention and the evaluation process (Domenech-Rodríguez & Wieling, 2004). Although each cultural adaptation model has its distinct characteristics, they all share a clear common trait, which is to underline the importance of cultural relevance and ensure that an intervention is in tune with the cultural values and daily experiences of target populations.

Systematic reviews and meta-analytic studies focused on the cultural adaptation of mental health interventions have been trying to evaluate the relevance of cultural adaptation. Some studies have suggested that culturally adapted interventions are more effective than interventions that do not pass through a process of cultural adaptation (e.g., Benish, Quintana, & Wampold, 2011; Sundell, Beelmann, Hasson, & Schwarz, 2016). The most recent meta-analysis supporting this hypothesis is from Soto, Smith, Griner, Domenech-Rodríguez, and Bernal (2018). These authors aimed to analyze the effectiveness of culturally adapted mental health interventions with different ethnic groups in the United States and Canada. Their results support that cultural adaptation did in fact produce better effectiveness levels than traditional treatments when implemented with different ethnic groups. However, the authors

acknowledge that results from this study should be interpreted with caution. Firstly, a key limitation found was that many studies did not thoroughly describe the adaptations made to the interventions, and thus they cannot infer on which adapted components were effective. Secondly, this meta-analysis only included studies that were conducted in the United States or Canada, thus not examining their effectiveness in other countries. For these reasons, the authors recommend that further research on the subject should be extended.

In contrast with Soto et al. (2018) study, the meta-analysis of Gardner et al. (2016) assessed the effectiveness of transporting evidence-based parenting interventions developed worldwide for childhood behavior problems and found that transporting interventions was effective, even when no cultural adaptation was made (Gardner et al., 2016). Moreover, one interesting result found was that interventions that were imported to more culturally different countries (e. g., Hong Kong) seemed to have more similar levels of effectiveness with the intervention's origin country (i.e., usually English-speaking countries) (Gardner et al., 2016). Despite these promising results, this study is not free from limitations. The authors point out that this meta-analysis did not include trials from the origin countries, and so a direct comparison of effectiveness levels between original and transported interventions was not analyzed. In addition, they argue that the inclusion of a small number of studies (i.e., 14 studies) lead to high levels of heterogeneity and low power (Gardner et al., 2016).

Taking into consideration the current rapid implementation and transportation of mental health interventions, it is important to highlight that most previous research focuses on mental health interventions in general. In fact, only Gardner et al. (2016) meta-analysis analyzed the effectiveness of transporting parenting interventions designed for childhood behavior problems. In this context, it is very important to broaden empirical evidence on this specific area of expertise. Therefore, the present meta-analysis intends to: 1) examine the effectiveness of parenting interventions developed in Anglosphere countries, when transported to non-Anglosphere countries, as well as compare effectiveness levels between Anglosphere and non-Anglosphere trials; and 2) perform an exploratory subgroup analysis regarding contextual and research factors that may impact the adaptation of the intervention, such as not employing cultural adaption methods, implementing the program in real-world setting and the use of waitlist/no intervention control groups.

This review aims to overcome some barriers and limitations found in the existing literature. On one hand, we only include studies of parenting interventions designed for childhood behavior problems. Thus, our results and conclusions do not come from generalizations made from other types of interventions. On the other hand, this review allows a comparison between the original and transported interventions, and thus interpretations regarding the effectiveness of transported interventions are done considering the effectiveness of the original interventions. Finally, to the best of our knowledge, this is the first meta-analysis that takes into account that most evidence-based parenting interventions that are being disseminated worldwide come from Anglosphere countries. This can have a major impact, since our results can better inform policy makers.

2. Methods

2.1. Eligibility criteria

This systematic review with meta-analysis consists of parenting interventions that: a) were created and developed in an Anglosphere setting; b) were tested in non-Anglosphere countries; c) their main aim focused on reducing childhood behavioral problems; and d) were designed for children ranging from two to 12 years old. This range of age (i.e., from two to 12 years old) was chosen as a criterion for this review for two main reasons. Firstly, research suggests that parenting interventions designed for treating childhood behavior problems can be more effective across children ranging from 2 to 11 years old (Gardner et al., 2018). Secondly, most well-known, evidence-based parenting interventions for childhood behavior problems were primarily designed for children between the ages of two and 12 years old.

Taking into account our criteria aforementioned, studies were included if: a) they had an experimental randomized trial design, where the control condition was either no treatment, a waiting list, care as usual or another intervention; b) children had some degree of identified behavioral problems; c) the mean age of children ($+ - 2$ SD) was comprised between two and 12 years old; d) parent training was the only component of intervention, and e) they incorporated at least one outcome measure related to childhood behavioral problems. The first Anglosphere randomized trial of each parenting intervention was included. This intended to avoid that a residual proportion of intervention protocols tested multiple times in high-income Anglosphere countries dominated the analyses, and also artificially increase the level of homogeneity between Anglosphere trials. It also made it possible to guarantee that the randomized trial testing the intervention protocol in its country of origin always preceded the intervention protocol that was transported to a non-Anglosphere context. In contrast, all non-Anglosphere randomized trials were included, in order to account for the diverse cultural settings, critical for the main purpose of this meta-analysis.

Furthermore, studies were excluded if: a) included multicomponent interventions (e.g., targeting parents, children and teachers, or at least two of these populations); c) >50% of the sample was composed by children with comorbid disorders, such as autism, attention deficit hyperactivity disorder (ADHD) or developmental disorders; d) >50% of the sample of families had specific familial characteristics (e.g., military families, divorced mothers, battered women) or significant clinical conditions (e.g., parents with a psychiatric disorder); e) there were discrepancies of sample characteristics (e.g., significant age differences) between trials of a specific parenting intervention; f) were not written in English, Portuguese or Spanish; g) only reported follow-up results; h) mildly modifications to parenting interventions were performed; and i) presented duplication of data.

2.2. Information sources and search strategies

Five electronic databases were used to search for published and unpublished relevant studies, namely: PsycINFO; Web of Science Core Collection; PubMed; SCOPUS and Google Scholar. In this online retrieving method, we applied a search equation including the following terms: ("parent* program*" OR "parent* intervention" OR "parent* training") AND (child*) AND ("externalizing behav*" OR "disruptive behav*" OR "behav* problem*" OR "conduct problem*" OR "conduct disorder*") AND ("randomized controlled" OR "randomised controlled" OR RCT). Forty-five combinations were created, which were limited-searched to title, abstract and keywords. Furthermore, we did not specify any time criterion.

Searches were conducted during December 2020, and an additional search was performed in September 2022 with the same search strategy to ensure data included was up to date. Hand-search and snowball methods were also used in both moments. Some of the authors were contacted to obtain more information about their studies. The search strategy was limited to data written in English, Portuguese or Spanish.

2.3. Data collection and analysis

Endnote software (version X7) was used to manage studies retrieved from literature searches. Subsequently to removing duplicates, two independent investigators assessed both titles, abstracts and full reports that appeared to meet our eligibility criteria. A third reviewer supervised the data collection process and assessed discrepancies. Intervention effects were analyzed if means and standard deviations (SD) were available for pre- and post-intervention data points.

Review Manager (RevMan, version 5.4) statistical software was used to conduct our meta-analysis, and complementary analysis of prediction intervals for between-study heterogeneity was performed using the Comprehensive Meta-Analysis Software (CMA).

A random-effects model was selected given the heterogeneity of included studies, namely in terms of intervention content, measures and other methodological factors (e.g., wide range of ages between samples). Standardized mean differences (SMD) and 95% confidence intervals (CI) were computed for individual studies, subgroup analyses and pooled estimates. Mean differences represent change scores between baseline and post-intervention data for each trial arm (Twisk et al., 2018), and are recommended to measure effect sizes when studies include small sample sizes (Higgins et al., 2021). Follow-up data was not included in this meta-analysis due to comparison groups being lost in the trials that used waitlist/no intervention conditions.

Heterogeneity was assessed using Q statistics, I^2 percentages and *p*-values. Significant Q test results indicate that between studies variance exists, and thus total variance is not equal to the within-studies variance. In addition, I^2 statistics completes the analysis of heterogeneity with an intuitive description of the variability in effect estimates that is due to heterogeneity rather than sampling error. In this context, 30% to 60% may represent moderate heterogeneity, 50% to 90% substantial heterogeneity, and 75% to 100% considerable heterogeneity (Higgins et al., 2021). A complementary analysis of between-study heterogeneity was assessed using 95% prediction intervals (PI). These provide essential information regarding the range of effect estimates of future studies (IntHout, Ioannidis, Rovers, & Goeman, 2016).

Further subgroup analyses were performed, when sufficient data were available, in order to assess to which extent the effect size varied depending on contextual and research factors, such as the employment of cultural adaptation methods, services offered to the control groups, and implementation settings (i.e., research conditions versus real-world conditions).

2.4. Risk of bias

In order to control risk of bias and to prevent inaccurate extrapolation of results, the methodological quality of studies was analyzed using the revised version of the Cochrane risk-of-bias tool for randomized trials (RoB 2.0) (Sterne et al., 2019), which is one of the most commonly used tools for randomized controlled trials (Higgins et al., 2021; Ma et al., 2020). The Cochrane risk-of-bias tool allows reviewers to assess five domains through which bias might be introduced into a trial, including randomization process bias, bias due to deviations from intended interventions, missing outcome data bias, bias in the measurement of the outcome, bias in the selection of the reported result and, at last, overall bias (Sterne et al., 2019). Furthermore, studies can be assessed as having "low", "some concerns" or "high" risk of bias in each domain and its overall domain (Sterne et al., 2019). According to the RoB 2.0 guidelines, assessing an individual domain as having a certain level of bias implies that the overall risk of bias is at least as severe (Higgins, Savović, Page, & Sterne, 2019).

No exclusion criteria regarding studies with high risk of bias was employed, because we predict that a significant risk of bias will be found throughout the studies included in our analysis. This is due to the fact that assessing risk of bias with the RoB 2.0 tool is dependent on consulting several reports regarding a specific trial (such as protocols, registrations or parallel publications), and since trials included are from non-Anglosphere countries, some reporting data may not be written in English, Spanish or Portuguese. Therefore, we predict that we will not have access to some information pertinent to analyze risk of bias of a certain trial, thus impacting its assessment. Furthermore, most experimental studies of psychological interventions conduct an intention-to-treat analysis (ITT), which by itself has a higher risk of bias inherent, due to the management of missing outcome data (Page, Altman, & Egger, 2022). Notwithstanding, a sensitivity analysis will be performed

by removing studies with two or more domains with high risk of bias, in order to assess its impact in our main results.

This review was structured in accordance with PRISMA Statement (Moher, Liberati, Tetzlaff, & Altman, 2009).

3. Results

3.1. Literature search

A total of 24,807 articles were found. After removing all duplicates and studies that were not written in English, Spanish or Portuguese, the search strategy identified 7820 articles that were deemed potentially eligible. After reviewing titles, abstracts and full text, 18 studies were included from the initial search. Additionally, two studies were manually integrated, namely Patterson, Chamberlain, and Reid (1982) and Schuhmann, Foote, Eyberg, Boggs, and Algina (1998). We hypothesize that these studies did not appear in our search strategy due to their date of publish or inadequate insertion in the databases. Nonetheless, they ended up being detected by snowball methods, and thus included in this review.

Studies (*k* = 7796) were mainly excluded for: a) not having an experimental design; b) interventions were not created in an Anglosphere country; c) children did not present behavioral problems; d) the mean age of children included (+ − 2 SD) did not range between two and 12 years old; e) interventions had not yet been implemented in a non-Anglosphere country; f) >50% of children from the samples had comorbid disorders associated; and g) >50% of families from the samples had specific sociocultural characteristics or clinical disorders. Additionally, Anglosphere studies that were not conducted in the country of origin where the parenting intervention was created and developed were also excluded.

3.2. Characteristics of included studies

A total of 20 trials were included in the meta-analysis (see Fig. 1), which consisted in an inclusion of six interventions, namely Incredible Years, Cope, Generation Parent Management Training – Oregon Model (GenerationPMTO), Parent-Child Interaction Therapy (PCIT), Triple P Discussion Groups (Level 3) and Group Triple P (Level 4). Similar content and theoretical backgrounds can be found in all six interventions. In addition, with the exception of PCIT and GenerationPMTO, interventions are underpinned in a group-based format and rely on a collaborative model to implement the program, usually using group leaders to facilitate the intervention. PCIT and GenerationPMTO base on a more flexible type of learning, where a therapist works with a family individually and a set of criteria have to be met in order to proceed with the next stage of the intervention (Niec, 2018; Scavenius et al., 2020).

All interventions were developed in Anglosphere countries, more specifically in the United States of America (USA), Canada and Australia. In addition, their dissemination trials took place in nine non-Anglosphere countries, namely Sweden, Finland, Norway, Netherlands, Portugal, Denmark, Iceland, Panama and Hong Kong. The majority of families were native to the country where trials took place and participants had, on average, a medium socioeconomic status (SES). In most studies, children were included either by referral from health or social services providers, or from parents seeking help due to their concern about the behavior of their child. Moreover, eight studies screened childhood behavior problems for inclusion using clinical cutoff scores on parent-reported questionnaires.

In this scope, all studies assessed behavioral problems in childhood using parent reports (e.g., Eyberg Child Behavior Inventory – ECBI, Child Behavior Checklist – CBCL) and/or observational data (e.g., Dyadic Parent-Child Interaction Coding System – DPICS), and a few also included teacher-reported data. Overall, data were provided by mothers and only six studies also included father-reported data. Most trials assessed other outcomes, such as parental mental health or child
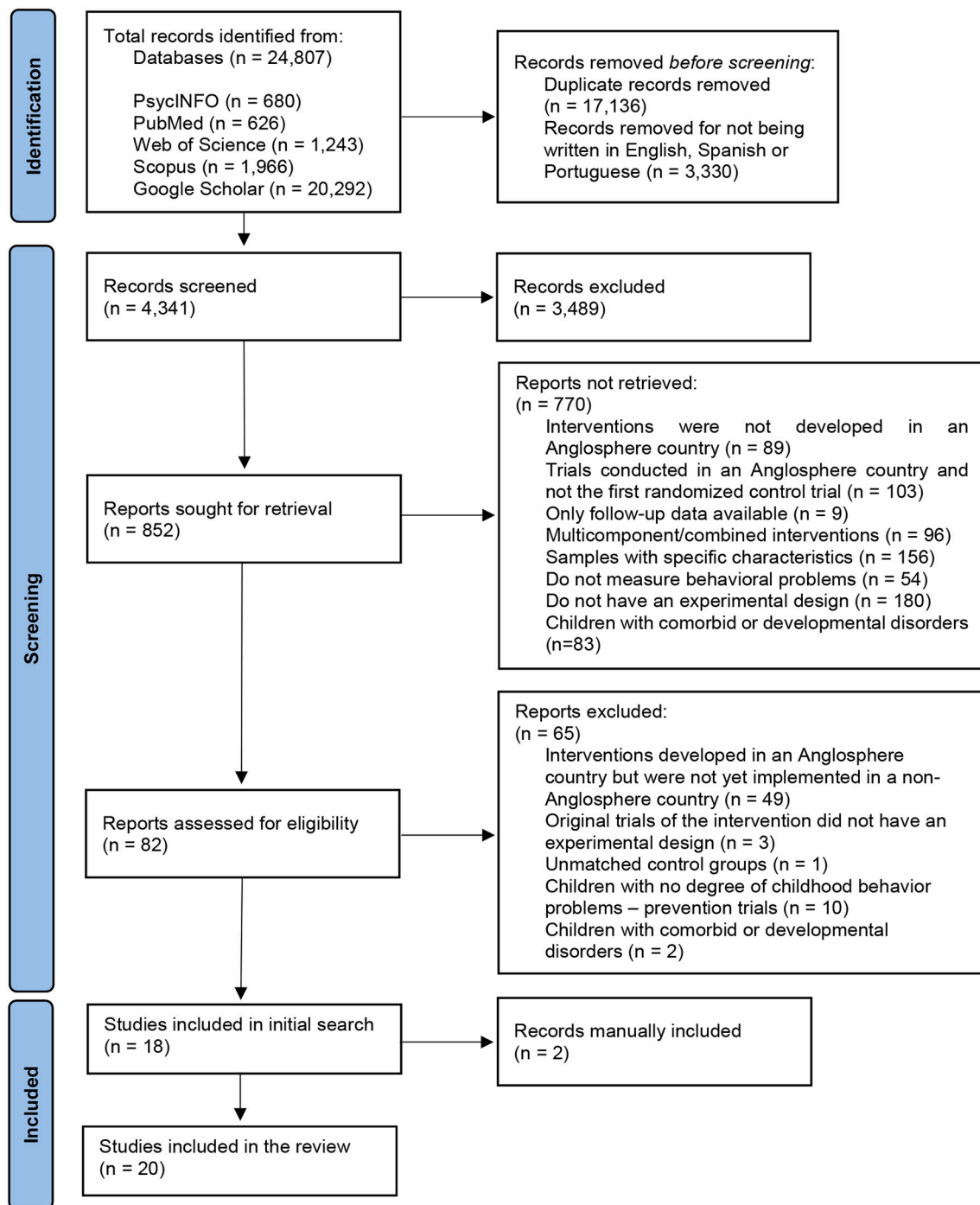
Fig. 1. Flow diagram for included studies.

internalizing behavior, which were not integrated in this study. Table 1 adds more information regarding included studies and interventions.

3.3. Effect of interventions (Anglosphere trials versus non-Anglosphere trials)

Change scores were based on the ECBI intensity scale, the Strengths and Difficulties Questionnaire (SDQ) externalizing scale, the Preschool and Kindergarten Behavior Scales (PKBS) – oppositional/explosive, observational data, and composite scores that included the ECBI intensity scale and four independent observations of aggression and inappropriate behavior at home. Table 2 shows a more detailed overview of the measures used to calculate change scores in each included

study. Overall, effect sizes under 0.20 were inferred to have insignificant indication of effect, those between −0.20 and − 0.40 as small effect sizes, −0.40 and − 0.75 as moderate effects and superior to −0.75 as large effect sizes (Cohen, 1988).

As shown in Fig. 2, six studies provided data for the subgroup of Anglosphere trials (i.e., efficacy trials of interventions developed in Anglosphere countries) and 14 provided data for the subgroup of non-Anglosphere trials (i.e., efficacy trials conducted in non-Anglosphere countries of interventions developed in Anglosphere countries).

Results suggest that more than three quarters (i.e., 79%) of non-Anglosphere trials showed statistically significant effect sizes, and that there were also no harmful effects across all the studied samples. Therefore, in both subgroups, results favored the intervention trial arm,

Table 1
Included interventions and countries of implementation.

Intervention	Country	Total Sample	Control/ Comparison Condition	Study Reference
Incredible Years	U.S.A. (c.o.)	159 families	No intervention	Webster-Stratton et al. (2004)
	Sweden	908 children	Waitlist	Stattin et al. (2015)
		62 mothers	Waitlist	Axberg and Broberg (2012)
	Finland	102 children	Services as usual	Karjalainen et al. (2019)
	Norway	127 children	Waitlist	Larsson et al. (2009)
	Netherlands	387 families	No intervention	Weeland et al. (2016)
		154 mothers	Waitlist	Leijten et al. (2017)
	Portugal	83 families	Waitlist	Homem (2014)
GenerationPMTO	U.S.A. (c.o.)	19 families	Services as usual	Patterson et al. (1982)
	Denmark	126 children	Services as usual	Scavenius et al. (2020)
	Norway	112 children	Services as usual	Ogden and Hagen (2008)
	Iceland	102 children	Services as usual	Sigmarsdóttir et al. (2014)
PCIT	U.S.A. (c.o.)	64 families	Waitlist	Schuhmann et al. (1998)
	Norway	81 children	Services as usual	Bjørseth and Wichstrøm (2016)
	Netherlands	45 children	Family Creative Therapy	Abrahamse et al. (2016)
Cope	Canada (c.o.)	150 children	No intervention	Cunningham and Boyle (1995)
	Sweden	908 children	Waitlist	Stattin et al. (2015)
Triple P Discussion Groups (Level 3)	Australia (c.o.)	85 parents	Waitlist	Dittman et al. (2016)
	Panama	108 families	No intervention	Meija et al. (2015)
Group Triple P (Level 4)	Australia (c.o.)	305 families	Waitlist	Sanders et al. (2000)
	Hong Kong	91 parents	Waitlist	Leung et al. (2003)

Note. c.o. – country of origin.

Table 2
Measures used to compute change scores.

Intervention	Study Reference	Measure
Incredible Years	Webster-Stratton et al. (2004)	Composite scores from ECBI: Intensity Scale and four independent observations of aggression and inappropriate behavior at home
	Stattin et al. (2015)	ECBI: Intensity Scale
	Axberg and Broberg (2012)	ECBI: Intensity Scale
	Karjalainen et al. (2019)	ECBI: Intensity Scale
	Larsson et al. (2009)	ECBI: Intensity Scale
	Weeland et al. (2016)	ECBI: Intensity Scale
	Leijten et al. (2017)	ECBI: Intensity Scale
	Homem (2014)	PKBS: Oppositional/Explosive Scale
GenerationPMTO	Patterson et al. (1982)	Observational data
	Scavenius et al. (2020)	SDQ: Externalizing Scale
	Ogden and Hagen (2008)	ECBI: Externalizing Scale
	Sigmarsdóttir et al. (2014)	ECBI: Externalizing Scale
PCIT	Schuhmann et al. (1998)	ECBI: Intensity Scale
	Bjørseth and Wichstrøm (2016)	ECBI: Intensity Scale
	Abrahamse et al. (2016)	ECBI: Intensity Scale
Cope	Cunningham and Boyle (1995)	Observational data
	Stattin et al. (2015)	ECBI: Intensity Scale
Triple P Discussion Groups (Level 3)	Dittman et al. (2016)	ECBI: Intensity Scale
	Meija et al. (2015)	ECBI: Intensity Scale
Group Triple P (Level 4)	Sanders et al. (2000)	ECBI: Intensity Scale
	Leung et al. (2003)	ECBI: Intensity Scale

Note. ECBI – Eyberg Child Behavior Inventory; PKBS - Preschool and Kindergarten Behavior Scales; SDQ - Strengths and Difficulties Questionnaire.

with Anglosphere trials (SMD = −0.81, 95% CI [−1.18, −0.45], $p <$ .0001) and non-Anglosphere trials (SMD = −0.50, 95% CI [−0.63, −0.37], $p <$ .00001) presenting a moderate effect. Overall results indicated significant moderate improvements to childhood behavioral problems (SMD = −0.57, 95% CI [−0.70, −0.44], $p <$ .00001), and subgroup differences were not significant, $\chi^2(1, N = 2529) = 2.52$, $p =$ .11. Tests showed substantial heterogeneity for Anglosphere trials, $Q(5) = 14.47$, $p <$ .01, $I^2 = 65\%$, 95% PI [−1.93, 0.31] and for overall analysis, $Q(20) = 43.18$, $p <$ .002, $I^2 = 54\%$, 95% PI [−1.01, −0.13], and moderate heterogeneity for non-Anglosphere trials, $Q(14) = 24.25$, $p <$ .05, $I^2 = 42\%$, 95% PI [−0.84, −0.16].

3.4. Factors associated with the effectiveness of a transported intervention

3.4.1. Interventions not culturally adapted

Exploratory analyses were conducted in order to test for factors that could attest for subgroup differences between Anglosphere and non-Anglosphere studies. Firstly, it was examined subgroup differences of interventions that were not targeted for cultural adaptation when transported to a non-Anglosphere country. In this context, any trial that

included interventions that had undergone any type of cultural adaptation, even if small ones (except for translation), was excluded. Therefore, a total of 12 studies were included in the analysis, five in the Anglosphere trials and seven in the non-Anglosphere trials subgroup (see Fig. 3). Overall results favored the intervention trials arm, in comparison with the control group, indicating moderate improvements to childhood behavioral problems that are statistically significant (SMD = −0.59, 95% CI [−0.76, −0.43], $p <$ .00001), and subgroup differences were not significant, $\chi^2(1, N = 1784) = 1.58$, $p =$ .21. Subgroup analysis showed that Anglosphere trials presented a moderate effect size (SMD = −0.78, 95% CI [−1.17, −0.39], $p =$ .0001), as well as the subgroup of non-Anglosphere trials effect (SMD = −0.51, 95% CI [−0.67, −0.34], $p <$ .00001). Tests for heterogeneity were significant for the subgroup of Anglosphere trials, $Q(4) = 13.61$, $p <$ .009, $I^2 = 71\%$, 95% PI [−2.13, 0.57] and for overall analysis, $Q(12) = 29.86$, $p =$ .003, $I^2 = 60\%$, 95% PI [−1.11, −0.07], but not for non-Anglosphere trials, $Q(7) = 13.04$, $p =$ .07, $I^2 = 46\%$, 95% PI [−0.92, −0.10].

3.4.2. Real-world setting

Subgroup differences of studies conducted in real-world settings, as defined by Wiesz, Donenberg, Han, and Weiss (1995), were examined. In this context, Anglosphere and non-Anglosphere subgroups included studies of interventions that were implemented under real-world conditions. A total of 13 studies were analyzed, three in the Anglosphere trials and 10 in the non-Anglosphere trials (see Fig. 4). The subgroup of

Pooled and subgroup analyses for anglosphere and non-anglosphere trials using change scores
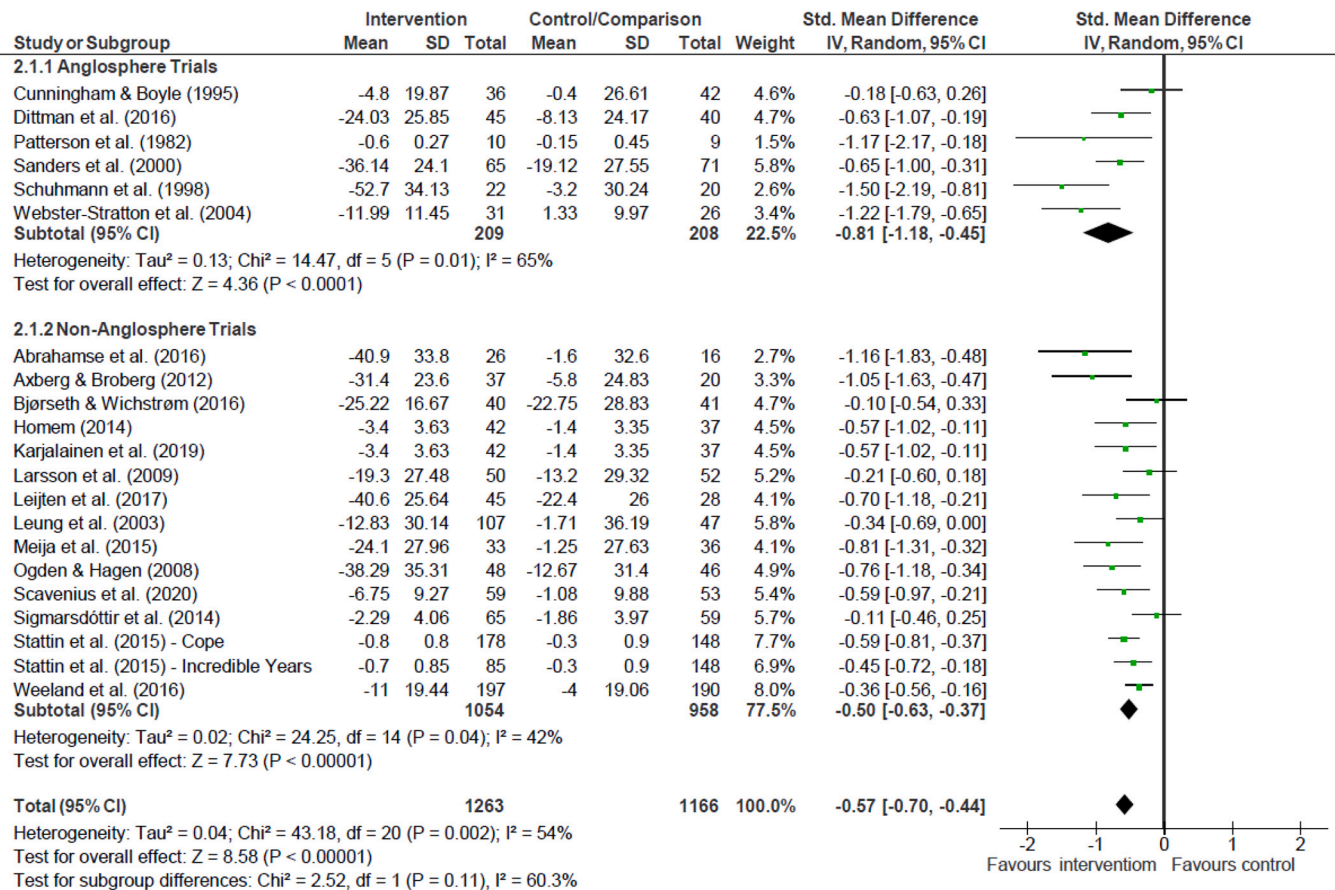
Study or Subgroup	Intervention Mean	SD	Total	Control/Comparison Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.1.1 Anglosphere Trials								
Cunningham & Boyle (1995)	-4.8	19.87	36	-0.4	26.61	42	4.6%	-0.18 [-0.63, 0.26]
Dittman et al. (2016)	-24.03	25.85	45	-8.13	24.17	40	4.7%	-0.63 [-1.07, -0.19]
Patterson et al. (1982)	-0.6	0.27	10	-0.15	0.45	9	1.5%	-1.17 [-2.17, -0.18]
Sanders et al. (2000)	-36.14	24.1	65	-19.12	27.55	71	5.8%	-0.65 [-1.00, -0.31]
Schuhmann et al. (1998)	-52.7	34.13	22	-3.2	30.24	20	2.6%	-1.50 [-2.19, -0.81]
Webster-Stratton et al. (2004)	-11.99	11.45	31	1.33	9.97	26	3.4%	-1.22 [-1.79, -0.65]
Subtotal (95% CI)			**209**			**208**	**22.5%**	**-0.81 [-1.18, -0.45]**
Heterogeneity: Tau² = 0.13; Chi² = 14.47, df = 5 (P = 0.01); I² = 65%								
Test for overall effect: Z = 4.36 (P < 0.0001)								
2.1.2 Non-Anglosphere Trials								
Abrahamse et al. (2016)	-40.9	33.8	26	-1.6	32.6	16	2.7%	-1.16 [-1.83, -0.48]
Axberg & Broberg (2012)	-31.4	23.6	37	-5.8	24.83	20	3.3%	-1.05 [-1.63, -0.47]
Bjørseth & Wichstrøm (2016)	-25.22	16.67	40	-22.75	28.83	41	4.7%	-0.10 [-0.54, 0.33]
Homem (2014)	-3.4	3.63	42	-1.4	3.35	37	4.5%	-0.57 [-1.02, -0.11]
Karjalainen et al. (2019)	-3.4	3.63	42	-1.4	3.35	37	4.5%	-0.57 [-1.02, -0.11]
Larsson et al. (2009)	-19.3	27.48	50	-13.2	29.32	52	5.2%	-0.21 [-0.60, 0.18]
Leijten et al. (2017)	-40.6	25.64	45	-22.4	26	28	4.1%	-0.70 [-1.18, -0.21]
Leung et al. (2003)	-12.83	30.14	107	-1.71	36.19	47	5.8%	-0.34 [-0.69, 0.00]
Meija et al. (2015)	-24.1	27.96	33	-1.25	27.63	36	4.1%	-0.81 [-1.31, -0.32]
Ogden & Hagen (2008)	-38.29	35.31	48	-12.67	31.4	46	4.9%	-0.76 [-1.18, -0.34]
Scavenius et al. (2020)	-6.75	9.27	59	-1.08	9.88	53	5.4%	-0.59 [-0.97, -0.21]
Sigmarsdóttir et al. (2014)	-2.29	4.06	65	-1.86	3.97	59	5.7%	-0.11 [-0.46, 0.25]
Stattin et al. (2015) - Cope	-0.8	0.8	178	-0.3	0.9	148	7.7%	-0.59 [-0.81, -0.37]
Stattin et al. (2015) - Incredible Years	-0.7	0.85	85	-0.3	0.9	148	6.9%	-0.45 [-0.72, -0.18]
Weeland et al. (2016)	-11	19.44	197	-4	19.06	190	8.0%	-0.36 [-0.56, -0.16]
Subtotal (95% CI)			**1054**			**958**	**77.5%**	**-0.50 [-0.63, -0.37]**
Heterogeneity: Tau² = 0.02; Chi² = 24.25, df = 14 (P = 0.04); I² = 42%								
Test for overall effect: Z = 7.73 (P < 0.00001)								
Total (95% CI)			**1263**			**1166**	**100.0%**	**-0.57 [-0.70, -0.44]**
Heterogeneity: Tau² = 0.04; Chi² = 43.18, df = 20 (P = 0.002); I² = 54%								
Test for overall effect: Z = 8.58 (P < 0.00001)								
Test for subgroup differences: Chi² = 2.52, df = 1 (P = 0.11), I² = 60.3%								

Std. Mean Difference IV, Random, 95% CI — -2 -1 0 1 2 — Favours intervention | Favours control

Fig. 2. Pooled and subgroup analyses for anglosphere and non-anglosphere trials using change scores.

Anglosphere trials presented a statistically significant large effect size (SMD = −0.91, 95% CI [−1.33, −0.50], $p < .0001$), therefore suggesting a clear preference for intervention over the control/comparison group. Moreover, statistically significant results were found for non-Anglosphere trials, which presented a moderate effect (SMD = 0.47, 95% CI [−0.68, −0.26], $p < .00001$), and for overall results, which indicated moderate improvements to childhood behavioral problems (SMD = 0.56, 95% CI [−0.76, −0.36], $p < .00001$). Subgroup differences were not significant, $\chi^2(1, N = 1207) = 3.42$, $p = .06$. In addition, tests for heterogeneity suggests substantial heterogeneity for non-Anglosphere trials, $Q(9) = 22.18$, $p < .008$, $I^2 = 59\%$, 95% PI [−1.09, 0.15], and for overall analysis, $Q(12) = 32.04$, $p < .001$, $I^2 = 63\%$, 95% PI [−1.22, 0.11]. Regarding the subgroup of Anglosphere trials, tests for heterogeneity suggests moderate heterogeneity, but results were not statistically significant, $Q(2) = 3.24$, $p < .20$, $I^2 = 38\%$, 95% PI [−4.80, 2.98].

3.4.3. Waitlist/no intervention control group condition

At last, it was examined subgroup differences of studies with waitlist/no intervention control groups. These include all studies where the control group had no type of intervention during the trial phase. Therefore, the analysis included 12 studies, four in the subgroup of Anglosphere trials, and their correspondent eight transported trials to non-Anglosphere countries, also with a waitlist/no intervention control group (see Fig. 5). Moderate effect sizes were found in the Anglosphere trials (SMD = −0.64, 95% CI [−1.00, −0.28], $p = .0004$), as well as in the non-Anglosphere (SMD = −0.54, 95% CI [−0.67, −0.41], $p < .00001$). In addition, subgroup differences were not significant, $\chi^2(1, N$

= 1828) = 0.26, $p = .61$. Tests for heterogeneity showed significant results for Anglosphere trials, $Q(3) = 7.91$, $p = .05$, $I^2 = 62\%$, 95% PI [−2.09, 0.81], but not significant results for non-Anglosphere trials, $Q(8) = 10.34$, $p = .24$, $I^2 = 23\%$, 95% PI [−0.82, −0.26] and for overall analysis, $Q(12) = 18.86$, $p = .09$, $I^2 = 36\%$, 95% PI [−0.91, −0.23].

3.5. Risk of bias

Using the RoB 2.0 tool (Table A1), assessment of risk of bias suggests that 14 studies (Abrahamse, Junger, van Wouwe, Boer, & Lindauer, 2016; Axberg & Broberg, 2012; Dittman, Farruggia, Keown, & Sanders, 2016; Homem, 2014; Karjalainen, Kiviruusu, Aronen, & Santalahti, 2019; Larsson et al., 2009; Leung, Sanders, Leung, Mak, & Lau, 2003; Meija, Calam, & Sanders, 2015; Ogden & Hagen, 2008; Patterson et al., 1982; Scavenius et al., 2020; Schuhmann et al., 1998; Stattin, Enebrink, Özdemir, & Giannotta, 2015; Webster-Stratton, Reid, & Hammond, 2004) have an overall high risk of bias and six have overall some concerns (Bjørseth & Wichstrøm, 2016; Cunningham & Boyle, 1995; Leijten, Raaijmakers, de Castro, van den Ban, & Matthys, 2017; Sigmarsdóttir et al. (2014); Sanders, Markie-Dadds, Tully, & Bor, 2000; Weeland et al., 2016). According to the analysis, missing outcome data is the domain of bias with lower risk throughout the studies, and the measurement of outcome is the domain with higher risk, due to the fact that most studies do not mention if outcome assessors were blinded to randomization and group conditions. In parallel, the selection of the reported result is the domain with more reporting gaps.

A sensitivity analysis was performed in order to assess the impact of bias in the main results of this meta-analysis. We identified two studies

Subgroup analysis: Non-anglosphere trials with no cultural adaptation compared with its correspondent anglosphere trial
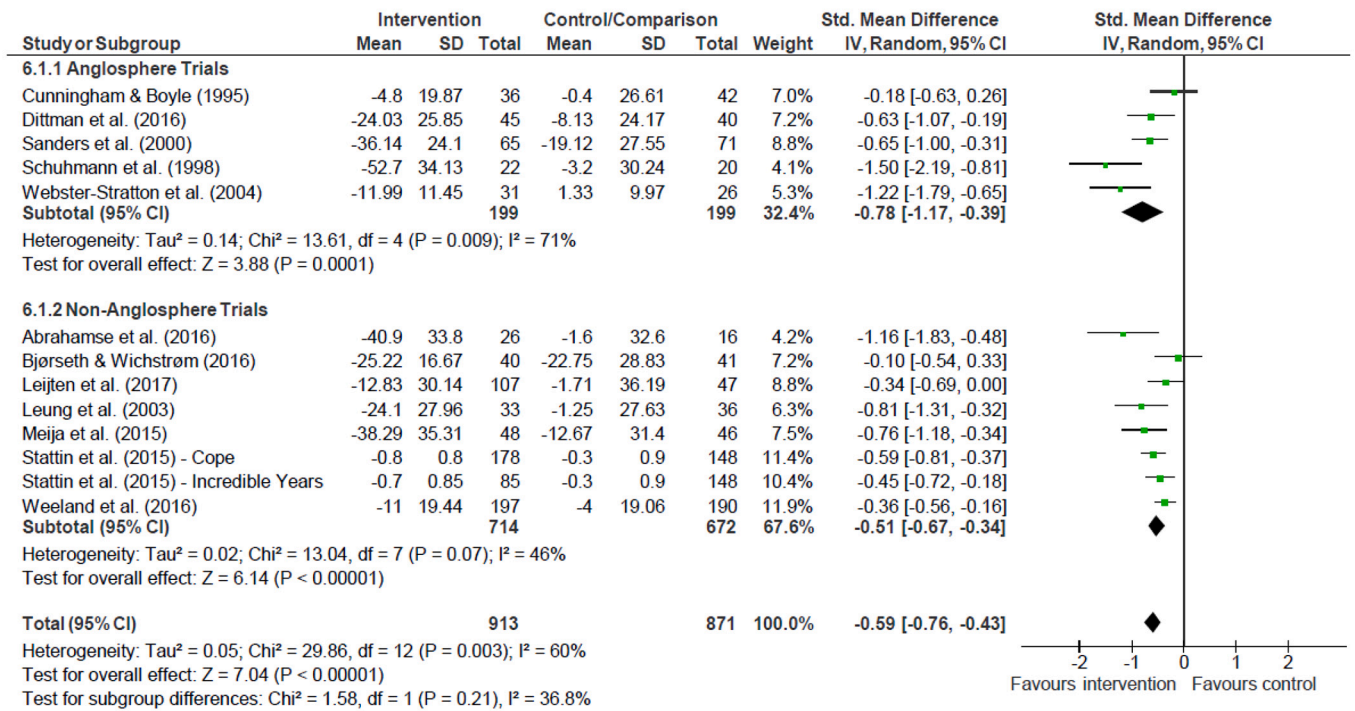
Study or Subgroup	Intervention Mean	SD	Total	Control/Comparison Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI	Std. Mean Difference IV, Random, 95% CI
6.1.1 Anglosphere Trials									
Cunningham & Boyle (1995)	-4.8	19.87	36	-0.4	26.61	42	7.0%	-0.18 [-0.63, 0.26]	
Dittman et al. (2016)	-24.03	25.85	45	-8.13	24.17	40	7.2%	-0.63 [-1.07, -0.19]	
Sanders et al. (2000)	-36.14	24.1	65	-19.12	27.55	71	8.8%	-0.65 [-1.00, -0.31]	
Schuhmann et al. (1998)	-52.7	34.13	22	-3.2	30.24	20	4.1%	-1.50 [-2.19, -0.81]	
Webster-Stratton et al. (2004)	-11.99	11.45	31	1.33	9.97	26	5.3%	-1.22 [-1.79, -0.65]	
Subtotal (95% CI)			199			199	32.4%	-0.78 [-1.17, -0.39]	
Heterogeneity: Tau² = 0.14; Chi² = 13.61, df = 4 (P = 0.009); I² = 71%									
Test for overall effect: Z = 3.88 (P = 0.0001)									
6.1.2 Non-Anglosphere Trials									
Abrahamse et al. (2016)	-40.9	33.8	26	-1.6	32.6	16	4.2%	-1.16 [-1.83, -0.48]	
Bjørseth & Wichstrøm (2016)	-25.22	16.67	40	-22.75	28.83	41	7.2%	-0.10 [-0.54, 0.33]	
Leijten et al. (2017)	-12.83	30.14	107	-1.71	36.19	47	8.8%	-0.34 [-0.69, 0.00]	
Leung et al. (2003)	-24.1	27.96	33	-1.25	27.63	36	6.3%	-0.81 [-1.31, -0.32]	
Meija et al. (2015)	-38.29	35.31	48	-12.67	31.4	46	7.5%	-0.76 [-1.18, -0.34]	
Stattin et al. (2015) - Cope	-0.8	0.8	178	-0.3	0.9	148	11.4%	-0.59 [-0.81, -0.37]	
Stattin et al. (2015) - Incredible Years	-0.7	0.85	85	-0.3	0.9	148	10.4%	-0.45 [-0.72, -0.18]	
Weeland et al. (2016)	-11	19.44	197	-4	19.06	190	11.9%	-0.36 [-0.56, -0.16]	
Subtotal (95% CI)			714			672	67.6%	-0.51 [-0.67, -0.34]	
Heterogeneity: Tau² = 0.02; Chi² = 13.04, df = 7 (P = 0.07); I² = 46%									
Test for overall effect: Z = 6.14 (P < 0.00001)									
Total (95% CI)			913			871	100.0%	-0.59 [-0.76, -0.43]	
Heterogeneity: Tau² = 0.05; Chi² = 29.86, df = 12 (P = 0.003); I² = 60%									
Test for overall effect: Z = 7.04 (P < 0.00001)									
Test for subgroup differences: Chi² = 1.58, df = 1 (P = 0.21), I² = 36.8%									

-2 -1 0 1 2
Favours intervention Favours control

Fig. 3. Subgroup analysis: Non-anglosphere trials with no cultural adaptation compared with its correspondent anglosphere trial.

Subgroup analysis: Real-world setting anglosphere trials versus *non-anglosphere trials*

Study or Subgroup	Intervention Mean	SD	Total	Control/Comparison Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI	Std. Mean Difference IV, Random, 95% CI
4.1.1 Anglosphere Trials									
Patterson et al. (1982)	-0.6	0.27	10	-0.15	0.45	9	3.1%	-1.17 [-2.17, -0.18]	
Sanders et al. (2000)	-36.14	24.1	65	-19.12	27.55	71	9.5%	-0.65 [-1.00, -0.31]	
Webster-Stratton et al. (2004)	-11.99	11.45	31	1.33	9.97	26	6.4%	-1.22 [-1.79, -0.65]	
Subtotal (95% CI)			106			106	19.0%	-0.91 [-1.33, -0.50]	
Heterogeneity: Tau² = 0.05; Chi² = 3.24, df = 2 (P = 0.20); I² = 38%									
Test for overall effect: Z = 4.29 (P < 0.0001)									
4.1.2 Non-Anglosphere Trials									
Abrahamse et al. (2016)	-40.9	33.8	26	-1.6	32.6	16	5.3%	-1.16 [-1.83, -0.48]	
Axberg & Broberg (2012)	-31.4	23.6	37	-5.8	24.83	20	6.3%	-1.05 [-1.63, -0.47]	
Bjørseth & Wichstrøm (2016)	-25.22	16.67	40	-22.75	28.83	41	8.1%	-0.10 [-0.54, 0.33]	
Karjalainen et al. (2019)	-19.3	27.48	50	-13.2	29.32	52	8.8%	-0.21 [-0.60, 0.18]	
Larsson et al. (2009)	-40.6	25.64	45	-22.4	26	28	7.4%	-0.70 [-1.18, -0.21]	
Leung et al. (2003)	-24.1	27.96	33	-1.25	27.63	36	7.3%	-0.81 [-1.31, -0.32]	
Ogden & Hagen (2008)	-6.75	9.27	59	-1.08	9.88	53	9.0%	-0.59 [-0.97, -0.21]	
Scavenius et al. (2020)	-2.29	4.06	65	-1.86	3.97	59	9.4%	-0.11 [-0.46, 0.25]	
Sigmarsdóttir et al. (2014)	-4.34	9.3	51	-3.32	8.5	51	8.8%	-0.11 [-0.50, 0.27]	
Stattin et al. (2015) - Incredible Years	-0.7	0.85	85	-0.3	0.9	148	10.6%	-0.45 [-0.72, -0.18]	
Subtotal (95% CI)			491			504	81.0%	-0.47 [-0.68, -0.26]	
Heterogeneity: Tau² = 0.06; Chi² = 22.18, df = 9 (P = 0.008); I² = 59%									
Test for overall effect: Z = 4.42 (P < 0.00001)									
Total (95% CI)			597			610	100.0%	-0.56 [-0.76, -0.36]	
Heterogeneity: Tau² = 0.08; Chi² = 32.04, df = 12 (P = 0.001); I² = 63%									
Test for overall effect: Z = 5.50 (P < 0.00001)									
Test for subgroup differences: Chi² = 3.42, df = 1 (P = 0.06), I² = 70.8%									

-2 -1 0 1 2
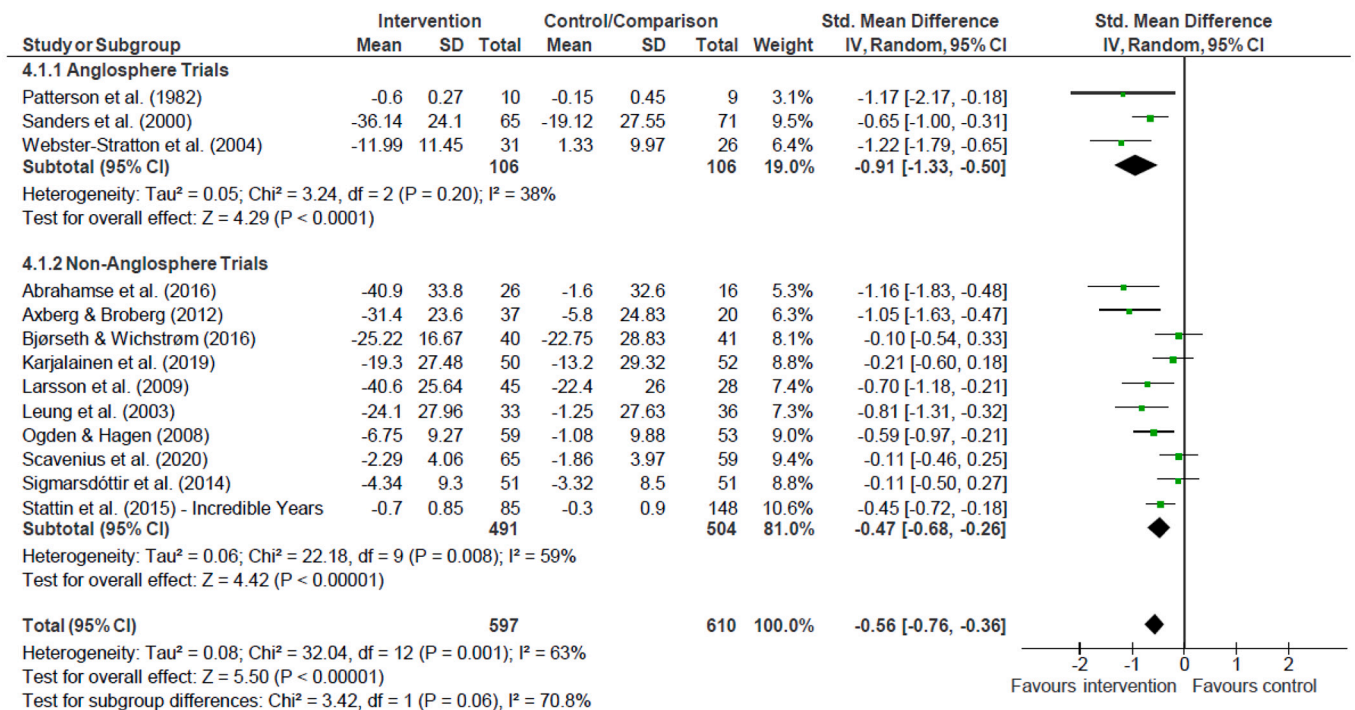Favours intervention Favours control

Fig. 4. Subgroup analysis: real-world setting anglosphere trials versus non-anglosphere trials.

that had two or more domains with high risk of bias, namely Abrahamse et al. (2016) and Homem (2014), and excluded them from the analysis. As shown in Fig. 6, results still favored the intervention trial arm, with Anglosphere trials (SMD = −0.81, 95% CI [−1.18, −0.45], $p < .0001$) and non-Anglosphere trials (SMD = −0.47, 95% CI [−0.60, −0.35], $p < .00001$) presenting a moderate effect. Overall results indicated significant moderate improvements to childhood behavioral problems (SMD = −0.55, 95% CI [−0.69, −0.42], $p < .00001$), and subgroup differences

Subgroup analysis: Anglosphere and non-anglosphere trials with a waitlist or no intervention control groups
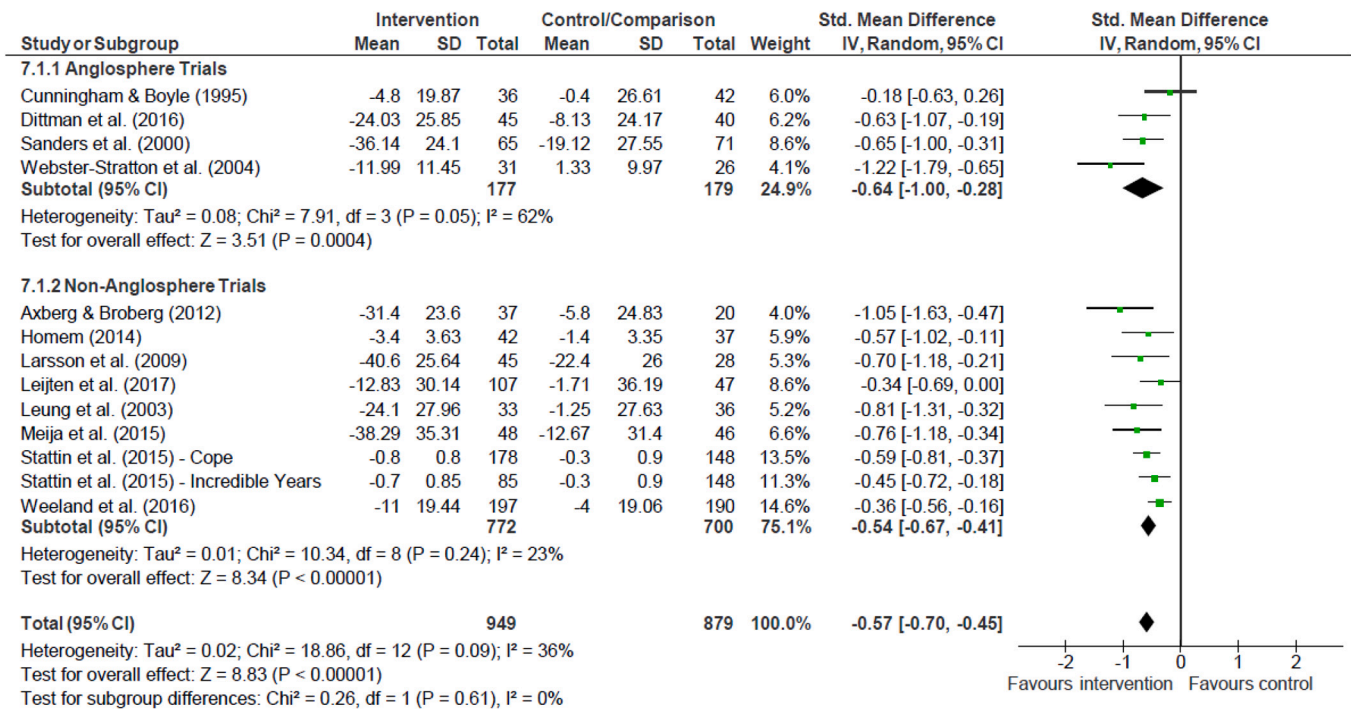
Study or Subgroup	Intervention Mean	SD	Total	Control/Comparison Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
7.1.1 Anglosphere Trials								
Cunningham & Boyle (1995)	-4.8	19.87	36	-0.4	26.61	42	6.0%	-0.18 [-0.63, 0.26]
Dittman et al. (2016)	-24.03	25.85	45	-8.13	24.17	40	6.2%	-0.63 [-1.07, -0.19]
Sanders et al. (2000)	-36.14	24.1	65	-19.12	27.55	71	8.6%	-0.65 [-1.00, -0.31]
Webster-Stratton et al. (2004)	-11.99	11.45	31	1.33	9.97	26	4.1%	-1.22 [-1.79, -0.65]
Subtotal (95% CI)			**177**			**179**	**24.9%**	**-0.64 [-1.00, -0.28]**
Heterogeneity: Tau² = 0.08; Chi² = 7.91, df = 3 (P = 0.05); I² = 62%								
Test for overall effect: Z = 3.51 (P = 0.0004)								
7.1.2 Non-Anglosphere Trials								
Axberg & Broberg (2012)	-31.4	23.6	37	-5.8	24.83	20	4.0%	-1.05 [-1.63, -0.47]
Homem (2014)	-3.4	3.63	42	-1.4	3.35	37	5.9%	-0.57 [-1.02, -0.11]
Larsson et al. (2009)	-40.6	25.64	45	-22.4	26	28	5.3%	-0.70 [-1.18, -0.21]
Leijten et al. (2017)	-12.83	30.14	107	-1.71	36.19	47	8.6%	-0.34 [-0.69, 0.00]
Leung et al. (2003)	-24.1	27.96	33	-1.25	27.63	36	5.2%	-0.81 [-1.31, -0.32]
Meija et al. (2015)	-38.29	35.31	48	-12.67	31.4	46	6.6%	-0.76 [-1.18, -0.34]
Stattin et al. (2015) - Cope	-0.8	0.8	178	-0.3	0.9	148	13.5%	-0.59 [-0.81, -0.37]
Stattin et al. (2015) - Incredible Years	-0.7	0.85	85	-0.3	0.9	148	11.3%	-0.45 [-0.72, -0.18]
Weeland et al. (2016)	-11	19.44	197	-4	19.06	190	14.6%	-0.36 [-0.56, -0.16]
Subtotal (95% CI)			**772**			**700**	**75.1%**	**-0.54 [-0.67, -0.41]**
Heterogeneity: Tau² = 0.01; Chi² = 10.34, df = 8 (P = 0.24); I² = 23%								
Test for overall effect: Z = 8.34 (P < 0.00001)								
Total (95% CI)			**949**			**879**	**100.0%**	**-0.57 [-0.70, -0.45]**
Heterogeneity: Tau² = 0.02; Chi² = 18.86, df = 12 (P = 0.09); I² = 36%								
Test for overall effect: Z = 8.83 (P < 0.00001)								
Test for subgroup differences: Chi² = 0.26, df = 1 (P = 0.61), I² = 0%								

Favours intervention ── Favours control

Fig. 5. Subgroup analysis: anglosphere and non-anglosphere trials with a waitlist or no intervention control groups.

Sensitivity analysis: Removal of studies with two or more domains with high risk of bias

Study or Subgroup	Intervention Mean	SD	Total	Control/Comparison Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
2.1.1 Anglosphere Trials								
Cunningham & Boyle (1995)	-4.8	19.87	36	-0.4	26.61	42	4.9%	-0.18 [-0.63, 0.26]
Dittman et al. (2016)	-24.03	25.85	45	-8.13	24.17	40	5.0%	-0.63 [-1.07, -0.19]
Patterson et al. (1982)	-0.6	0.27	10	-0.15	0.45	9	1.6%	-1.17 [-2.17, -0.18]
Sanders et al. (2000)	-36.14	24.1	65	-19.12	27.55	71	6.3%	-0.65 [-1.00, -0.31]
Schuhmann et al. (1998)	-52.7	34.13	22	-3.2	30.24	20	2.8%	-1.50 [-2.19, -0.81]
Webster-Stratton et al. (2004)	-11.99	11.45	31	1.33	9.97	26	3.7%	-1.22 [-1.79, -0.65]
Subtotal (95% CI)			**209**			**208**	**24.2%**	**-0.81 [-1.18, -0.45]**
Heterogeneity: Tau² = 0.13; Chi² = 14.47, df = 5 (P = 0.01); I² = 65%								
Test for overall effect: Z = 4.36 (P < 0.0001)								
2.1.2 Non-Anglosphere Trials								
Axberg & Broberg (2012)	-31.4	23.6	37	-5.8	24.83	20	3.6%	-1.05 [-1.63, -0.47]
Bjørseth & Wichstrøm (2016)	-25.22	16.67	40	-22.75	28.83	41	5.0%	-0.10 [-0.54, 0.33]
Karjalainen et al. (2019)	-3.4	3.63	42	-1.4	3.35	37	4.9%	-0.57 [-1.02, -0.11]
Larsson et al. (2009)	-19.3	27.48	50	-13.2	29.32	52	5.6%	-0.21 [-0.60, 0.18]
Leijten et al. (2017)	-40.6	25.64	45	-22.4	26	28	4.5%	-0.70 [-1.18, -0.21]
Leung et al. (2003)	-12.83	30.14	107	-1.71	36.19	47	6.3%	-0.34 [-0.69, 0.00]
Meija et al. (2015)	-24.1	27.96	33	-1.25	27.63	36	4.4%	-0.81 [-1.31, -0.32]
Ogden & Hagen (2008)	-38.29	35.31	48	-12.67	31.4	46	5.2%	-0.76 [-1.18, -0.34]
Scavenius et al. (2020)	-6.75	9.27	59	-1.08	9.88	53	5.8%	-0.59 [-0.97, -0.21]
Sigmarsdóttir et al. (2014)	-2.29	4.06	65	-1.86	3.97	59	6.2%	-0.11 [-0.46, 0.25]
Stattin et al. (2015) - Cope	-0.8	0.8	178	-0.3	0.9	148	8.3%	-0.59 [-0.81, -0.37]
Stattin et al. (2015) - Incredible Years	-0.7	0.85	85	-0.3	0.9	148	7.5%	-0.45 [-0.72, -0.18]
Weeland et al. (2016)	-11	19.44	197	-4	19.06	190	8.6%	-0.36 [-0.56, -0.16]
Subtotal (95% CI)			**986**			**905**	**75.8%**	**-0.47 [-0.60, -0.35]**
Heterogeneity: Tau² = 0.02; Chi² = 20.10, df = 12 (P = 0.07); I² = 40%								
Test for overall effect: Z = 7.27 (P < 0.00001)								
Total (95% CI)			**1195**			**1113**	**100.0%**	**-0.55 [-0.69, -0.42]**
Heterogeneity: Tau² = 0.04; Chi² = 39.62, df = 18 (P = 0.002); I² = 55%								
Test for overall effect: Z = 8.04 (P < 0.00001)								
Test for subgroup differences: Chi² = 2.95, df = 1 (P = 0.09), I² = 66.2%								

Risk of Bias: A B C D E F G

Favours interventiom ── Favours control

Fig. 6. Sensitivity analysis: removal of studies with two or more domains with high risk of bias.

were still not significant, $\chi^2(1, N = 2308) = 2.95, p = .09$. Tests showed substantial heterogeneity for Anglosphere trials, $Q(5) = 14.47, p < .01$, $I^2 = 65\%$, 95% PI $[-1.93, 0.31]$ and for overall analysis, $Q(18) = 39.62$, $p < .002, I^2 = 55\%$, 95% PI $[-0.99, -0.11]$, and moderate heterogeneity for non-Anglosphere trials, although results were not statistically significant, $Q(12) = 20.10, p = .07, I^2 = 40\%$, 95% PI $[-0.81, -0.13]$.

4. Discussion

This meta-analysis examined the clinical effectiveness of parenting interventions developed in Anglosphere countries, when transported to non-Anglosphere countries. Moreover, this study also assessed the impact of specific research and context factors on cross-cultural transportation, such as the degree of cultural adaptation, the role of real-world settings and choice of control group. It included research undertaken with six well-known parenting interventions, namely Incredible Years, GenerationPMTO, PCIT, Cope, Triple P Level 3 and Level 4. These parenting interventions are well-established, manualized interventions, mostly including training and certification methods to guarantee fidelity. In the end, 20 experimental studies were assessed, including six studies undertaken in Anglosphere countries, namely in the country where the intervention was created, and 14 dissemination trials conducted in nine different non-Anglosphere countries, each of which met criteria to be classified as high income, developed countries (United Nations Development Programme, 2020).

Studies included in this meta-analysis are of relevance to the scientific literature. Firstly, there is a global recognition of the need to deliver parenting interventions for the prevention and treatment of childhood behavior problems, and thus analyzing their impact is of foremost importance. In the same sense, there are communities who experience difficulties accessing evidence-based parenting interventions, and thus it is important to disseminate and expand the access to interventions. Testing parenting programs, designed for childhood behavioral problems, in the community, produces evidence that will not only better inform policy makers, but also raise awareness among professionals and parents. Moreover, transporting and analyzing parenting interventions of Anglosphere backgrounds in non-Anglosphere countries expands the literature in regard to the impact of culture on intervention's content and implementation procedures, which allows researchers to draw suggestions and guidelines for future research in this field.

Regarding our results, findings from our primary analysis suggest that these parenting interventions can indeed be successfully transported to non-Anglosphere countries, with rather different cultural backgrounds, and potentially maintain effectiveness. Meta-analytic results showed moderate mean effect sizes for non-Anglosphere trials. This represented a decrease in the mean effect size when compared to Anglosphere trials, but this difference was not statistically significant. Furthermore, 95% prediction intervals suggest that there is less than a 5% chance that a future implementation of these parenting interventions, under these conditions, will not be effective in reducing behavioral problems. This result supports the conclusion that Anglosphere parenting interventions, when transported to non-Anglosphere contexts, are effective, and therefore can help by providing parents with useful strategies to deal with childhood behavior problems.

In addition, the effects achieved after transportation to non-Anglosphere countries may suggest that intervention principles, content and intervention methods of programs are robust and may not be affected by cultural differences from one cultural context to another. For example, techniques common to many evidence-based parenting programs, such as social learning theory and behavioral strategies, positive family communication and relationships, and child-led play, are universal, transcultural characteristics of developmentally effective parenting. The transcultural performance of these parenting programs may be assisted by features such as manualization, training, supervision and fidelity monitoring. It should also be noted that the non-Anglosphere domains involved in the dissemination trials were most frequently high income, Northern European countries with potentially lower levels of contrasting cultural differences in parenting practices and child development expectations.

Cultural adaptation is potentially a tool to achieve higher levels of effectiveness when transporting evidence-based interventions to new, distinct cultural contexts. The meta-analysis of Sundell et al. (2016) supported this hypothesis and their results showed that interventions culturally adapted had higher effect sizes when compared to those that were not culturally adapted. In contrast, results from this meta-analysis showed no significant differences between the original trials conducted in the Anglosphere countries and the non-Anglosphere trials of interventions that were not culturally adapted when transported. These findings present a more amenable perspective regarding the dissemination of interventions and suggest that substantial cultural adaptation may not be crucial in order to ensure effectiveness, even in the presence of cultural disparities.

Differences in effect sizes between Anglosphere trials and non-Anglosphere trials conducted in real-world settings were also found, but were not statistically significant. Though real-world conditions can be problematic to operationalize, this finding is consistent with previous findings (e.g., Michelson, Davenport, Dretzke, Barlow, & Day, 2013) that evidence-based parenting interventions can be effective when delivered under real-world conditions. Therefore, interventions with promising effect sizes in a given real-world setting from their country of origin can maintain a significant level of effectiveness in real-world conditions of a different country, as long as the given organization is set up to receive the program. Hence, our findings seem to support once more the importance of maintaining fidelity in order to guarantee its success of dissemination.

Finally, effect sizes were assessed on trials that used a waitlist/no intervention control group, as well as compared differences between subgroups. Interestingly, the overall effect of trials that used a waitlist/no intervention control group was very similar to our primary analysis that included all the trials. This finding suggests that, in this analysis, having a waitlist/no intervention control group did not potentially overestimate effect sizes. This topic has been of interest in the empirical evidence, with findings supporting that a waiting list or a no intervention control arm may not be the best approach due to their overestimation in effect levels (e.g., Cunningham, Kypri, & McCambridge, 2013). However, results from this study do seem to contradict this hypothesis. Moreover, moderate effect sizes were still observed in both subgroups, which indicates that excluding trials with a service as usual trial arm, does not substantially impact effect levels.

Notwithstanding the importance of these findings, this study is not free of limitations. Firstly, this meta-analysis is confined to only 20 studies and results are limited by the substantial heterogeneity present in almost all our analyses. Possible explanations for high heterogeneity focus on clinical diversity, meaning that effects may have been affected by variables that were not captured by the inclusion criteria, such as individual variables (e.g., severity of childhood behavior problems), family variables (e.g., ethnicity, socioeconomic status) or intervention variables (e.g., duration of intervention). Hence, subgroup comparisons should be seen as exploratory and caution in interpreting these results is recommended.

Secondly, the literature search conducted accordingly to our inclusion criteria did not find any experimental trials conducted in developing or underdeveloped countries, which may suggest that these interventions either have not been tested in those countries or were not rigorously done so. According to several authors (e.g., Acharya & Pathak, 2019; Rojas, Martínez, Martínez, Franco, & Jiménez-Molina, 2019), this may be due to limited resources to support research projects, lack of research personnel or absence of ethical review committees. Therefore, it is not possible to draw conclusions concerning the transportation of Anglosphere interventions to underdeveloped or developing countries. Thus, future experimental trials regarding transportation of evidence-based, Anglosphere interventions to these

countries should be conducted.

Thirdly, although parent-reported questionnaires were used to guarantee outcome comparison between studies, these were not blinded to participants' condition. In addition, there is a lack of data concerning violations in measurement invariance. Consequently, there is no guarantee that families conceptualized childhood behavior problems in a similar manner. Moreover, although criteria were established in order to define constructs such as real-world conditions and cultural adaptation, there were difficulties associated to categorizing studies into these terms. This was in general due to missing data concerning the modifications applied to the intervention and its manual, as well as concerning the setting in which the interventions were being implemented. Additionally, in regards to cultural adaptation, although some interventions may have not been considered as culturally adapted, there might have been inbuilt flexibilities in their manuals that allowed group leaders to adapt some of the content, thus complicating the analysis of cultural adaptation.

Finally, trials included in this meta-analysis had in general a relevant risk of bias, that was mainly due to reporting gaps. As mentioned before, the RoB 2.0 tool is dependent on accessing several reports regarding a specific trial (such as protocols, registrations or parallel publications), and since we include trials from non-Anglosphere countries, some reporting data were not available in English, Spanish or Portuguese. Therefore, there were reporting gaps in the assessment of risk of bias. Moreover, a significant number of studies included were dated before 2005, and thus study protocols and registration protocols were not available. We explored the effect of removing studies where a high risk of bias was present in two or more domains and generally found a similar pattern of results, thus supporting the conclusion that risk of bias had little influence on our primary conclusions. Notwithstanding, the amount of risk of bias found should be taken into account in the interpretation of our findings. Further studies may need to include a more detailed description of methodology and publish other reports of the trial in the same language as the trial study.

To the best of our knowledge, this is the first systematic review and meta-analysis that compares effectiveness levels of interventions developed in Anglosphere countries and their dissemination trials conducted in non-Anglosphere countries. Findings from this study add to the literature and support the thesis that transporting interventions that were designed in an Anglosphere context can be clinically effective when implemented in non-Anglosphere settings. Although this meta-analysis only included trials from developed countries, and therefore its findings may not be applicable to underdeveloped or developing

contexts, studies included were from many world regions with distinct cultural and societal backgrounds.

These findings take a step further and suggest that interventions not only may not need to go through an extensive process of cultural adaptation when transported to non-Anglosphere countries of the types included in the meta-analysis, but also can be implemented in real-world settings without losing substantial effect levels. Therefore, and given that interventions included were evidence-based and manualized, it is possible to hypothesize that as long as fidelity is maintained, significant effect sizes can be upheld in non-Anglosphere contexts. Moreover, it is encouraging verifying in this analysis that using a waitlist/no intervention control arm did not overestimate effect sizes significantly, meaning that using this type of control group still is potentially an efficient approach when designing a randomized controlled trial.

In conclusion, this study is a relevant contribution to the empirical evidence regarding cross-cultural transportability of parenting interventions for childhood behavior problems. Therefore, it is part of a relevant field of research that can impact the standpoint of policy makers and practitioners worldwide.

Contributors

Authors Laura Maciel, Crispin Day and Miguel Basto-Pereira designed and wrote the protocol. Authors Laura Maciel and Aitana Gomis-Pomares conducted literature searches and screening of studies. Author Laura Maciel conducted the statistical analysis and wrote the first draft of the manuscript. All authors contributed to and have approved the final manuscript.

Funding sources

Declaration of Competing Interest

There are no conflicts of interest by any author.

Acknowledgments

Appendix A. Appendix

Table A1
Risk of bias assessment according to RoB 2.0.

Study reference	Intervention	Weight (%)	D1	D2	D3	D4	D5	Overall
Abrahamse et al. (2016)	PCIT	3.1	+	−	+	−	!	−
Axberg and Broberg (2012)	Incredible Years	3.7	!	+	+	−	!	−
Bjørseth and Wichstrøm (2016)	PCIT	4.8	+	+	+	+	!	!
Cunningham and Boyle (1995)	Cope	4.7	+	!	+	+	!	!
Dittman et al. (2016)	Triple P Discussion Groups (Level 3)	4.8	!	!	+	−	!	−
Homem (2014)	Incredible Years	4.7	!	!	−	−	!	−
Karjalainen et al. (2019)	Incredible Years	4.7	!	!	+	−	!	−
Larsson et al. (2009)	Incredible Years	5.2	!	!	+	−	!	−
Leijten et al. (2017)	Incredible Years	4.4	+	!	+	!	!	!
Leung et al. (2003)	Group Triple P (Level 4)	5.6	!	!	+	−	!	−
Meija et al. (2015)	Triple P Discussion Groups (Level 3)	4.3	!	!	−	+	!	−
Ogden and Hagen (2008)	GenerationPMTO	4.9	+	!	+	−	!	−
Patterson et al. (1982)	GenerationPMTO	1.9	!	−	−	−	!	−
Scavenius et al. (2020)	GenerationPMTO	5.3	+	−	−	−	!	−
Schuhmann et al. (1998)	PCIT	3	!	−	+	−	!	−
Sigmarsdóttir et al. (2014)	GenerationPMTO	5.5	+	!	+	+	!	!

Table A1 (*continued*)

Study reference	Intervention	Weight (%)	D1	D2	D3	D4	D5	Overall
Stattin et al. (2015)	Cope	12.8	!	!	+	−	!	−
	Incredible Years							
Sanders et al. (2000)	Group Triple P (Level 4)	3.5	+	+	+	!	!	!
Webster-Stratton et al. (2004)	Incredible Years	3.8	!	−	+	+	!	−
Weeland et al. (2016)	Incredible Years	8	+	+	+	+	+	+

Note. D1 – Randomization process; D2 – Deviations from the intended interventions; D3 – Missing outcome data; D4 – Measurement of the outcome; D5 – Selection of the reported result; + Low risk;! Some concerns; − High risk.

References

Abrahamse, M. E., Junger, M., van Wouwe, M. A. M. M., Boer, F., & Lindauer, R. J. L. (2016). Treating child disruptive behavior in high-risk families: A comparative effectiveness trial from a community-based implementation. *Journal of Child and Family Studies, 25*, 1605–1622. https://doi.org/10.1007/s10826-015-0322-4

Acharya, K. P., & Pathak, S. (2019). Applied research in low-income countries: Why and how? *Frontiers in Research Metrics and Analytics, 4*(3), 1–9. https://doi.org/10.3389/frma.2019.00003

Axberg, U., & Broberg, A. G. (2012). Evaluation of "the incredible years" in Sweden: The transferability of an American parent-training program to Sweden. *Scandinavian Journal of Psychology, 53*(3), 224–232. https://doi.org/10.1111/j.1467-9450.2012.00955.x

Barlow, J., & Coren, E. (2017). The effectiveness of parenting programs: A review of Campbell reviews. *Research on Social Work Practice, 28*(1), 99–102. https://doi.org/10.1177/1049731517725184

Basto-Pereira, M., & Farrington, D. P. (2022). Developmental predictors of offending and persistence in crime: A systematic review of meta-analyses. *Aggression and Violent Behavior, 65*. https://doi.org/10.1016/j.avb.2022.101761

Baumann, A. A., Powell, B. J., Kohl, P. L., Penalba, V., Proctor, E. K., … Cabassa, L. J. (2015). Cultural adaptation and implementation of evidence-based parent-training: A systematic review and critique of guiding evidence. *Children and Youth Services Review, 53*, 113–120. https://doi.org/10.1016/j.childyouth.2015.03.025

Benish, S. G., Quintana, S., & Wampold, B. E. (2011). Culturally adapted psychotherapy and the legitimacy of myth: A direct-comparison meta-analysis. *Journal of Counselling Psychology, 58*(3), 279–289. https://doi.org/10.1037/a0023626

Bernal, G., Bonilla, J., & Bellido, C. (1995). Ecological validity and cultural sensitivity for outcome research: Issues for the cultural adaptation and development of psychosocial treatments with Hispanics. *Journal of Abnormal Child Psychology, 23*(1), 67–82. https://doi.org/10.1007/BF01447045

Bjørseth, Å., & Wichstrøm, L. (2016). Effectiveness of parent-child interaction therapy (PCIT) in the treatment of young children's behavior problems. A randomized controlled study. *PLoS One, 11*(9), Article e0159845. https://doi.org/10.1371/journal.pone.0159845

Castro, F. G., Barrera, M., & Holleran Steiker, L. K. (2010). Issues and challenges in the design of culturally adapted evidence-based interventions. *Annual Review of Clinical Psychology, 6*, 213–239.

Cheon, B. K., Melani, I., & Hong, Y. (2020). How USA-centric is psychology? An archival study of implicit assumptions of generalizability of findings to human nature based on origins of study samples. *Social Psychological and Personality Science, 11*(7), 928–937. https://doi.org/10.1177/1948550620927269

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associated, Publishers.

Cooper, D. K., Wieling, E., Domenech-Rodríguez, M. M., Garcia-Huidobro, D., Baumann, A., Meija, A., … Acevedo-Polakovich, I. D. (2019). Latinx mental health Scholars' experiences with cultural adaptation and implementation of systemic family interventions. *Family Process, 59*(2), 492–508. https://doi.org/10.1111/famp.12433

Cunningham, C. E., & Boyle, M. (1995). Large group community-based parenting programs for families of preschoolers at risk for disruptive behaviour disorders: utilization, cost-effectiveness, and outcome. *Journal of Child Psychology and Psychiatry, 36*(7), 1141–1159. https://doi.org/10.1111/j.1469-7610.1995.tb01362.x

Cunningham, J. A., Kypri, K., & McCambridge, J. (2013). Exploratory randomized controlled trial evaluating the impact of a waiting list control design. *BMC Medical Research Methodology, 13*(1), 150–157. https://doi.org/10.1186/1471-2288-13-150

Dittman, C. K., Farruggia, S. P., Keown, L. J., & Sanders, M. R. (2016). Dealing with disobedience: An evaluation of a brief parenting intervention for young children showing noncompliant behavior problems. *Child Psychiatry and Human Development, 47*(1), 102–112. https://doi.org/10.1007/s10578-015-0548-9

Domenech-Rodríguez, M., & Wieling, E. (2004). Developing culturally appropriate evidence-based treatments for interventions with ethnic minority populations. In M. Rastogi, & E. Wieling (Eds.), *Voices of color: First person accounts of ethnic minority therapists* (pp. 313–333). Thousand Oaks, CA: Sage Publications.

Elliot, D. S., & Mihalic, S. (2004). Issues in disseminating and replicating effective prevention programs. *Prevention Science, 5*(1), 47–53. https://doi.org/10.1023/b:prev.0000013981.28071.52

Eyberg, S. M., Boggs, S. R., & Algina, J. (1995). Parent-child interaction therapy: A psychosocial model for the treatment of young children with conduct problem behavior and their families. *Psychopharmacology Bulletin, 31*(1), 83–91.

Farrington, D. P., Gaffney, H., Lösel, F., & Ttofi, M. M. (2017). Systematic reviews of the effectiveness of developmental prevention programs in reducing delinquency, aggression, and bullying. *Aggression and Violent Behavior, 33*, 91–106. https://doi.org/10.1016/j.avb.2016.11.003

Florean, I. S., Dobrean, A., Pasarelu, C. R., Georgescu, R. D., & Milea, I. (2020). The efficacy of internet-based parenting programs for children and adolescents with behavior problems: A meta-analysis of randomized clinical trials. *Clinical Child and Family Psychology Review, 23*, 510–528. https://doi.org/10.1007/s10567-020-00326-0

Fong, V. L. (2010). Parent-Child Communication Problems and the Perceived Inadequacies of Chinese Only Children. In R. A. LeVine (Ed.), *Psychological Anthropology: A Reader on Self in Culture* (pp. 220–238). Wiley-Blackwell.

Gardner, F., Leijten, P., Melendez-Torres, G. J., Landau, S., Mann, J., Beecham, J., & Scott, S. (2018). The earlier the better? Individual participant data and traditional meta-analysis of age effects of parenting interventions. *Child Development, 90*(1), 7–19. https://doi.org/10.1111/cdev.13138

Gardner, F., Montgomery, P., & Knerr, W. (2016). Transporting evidence-based parenting programs for child problem behavior (age 3-10) between countries: Systematic review and meta-analysis. *Journal of Clinical Child & Adolescent Psychology, 45*(6), 749–762. https://doi.org/10.1080/15374416.2015.1015134

General Assembly Resolution 45/112. (1990). United Nations guidelines for the prevention of juvenile delinquency (Riyadh guidelines). https://undocs.org/pdf?symbol=en/A/RES/45/112.

Gonzales, N. A. (2017). Expanding the cultural adaptation framework for population-level impact. *Prevention Science, 18*(6), 689–693. https://doi.org/10.1007/s11121-017-0808-y

Heim, E., & Kohrt, B. A. (2019). Cultural adaptation of scalable psychological interventions: A new conceptual framework. *Clinical Psychology in Europe, 1*(4), 1–22. https://doi.org/10.32872/cpe.v1i4.37679

Henrich, J. (2020). *The WEIRDest people in the world: How the west became psychologically peculiar and particularly prosperous*. Farrar, Straus and Giroux.

Higgins, J. P. T., Savović, J., Page, M. J., & Sterne, A. C. (2019). Revised Cochrane risk-of-bias tool for randomized trials (RoB 2). In *Risk of bias tools: Current version of RoB 2*. https://drive.google.com/file/d/19R9savfPdCHC8XLz2iiMvL_71lPJERWK/view.

Higgins, J. P. T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J., & Welch, V. A. (2021). *Cochrane handbook for systematic reviews of interventions (version 6.2)*. Cochrane. https://training.cochrane.org/handbook.

Hill, N. E., & Tyson, D. F. (2008). Excavating culture: Ethnicity and context as predictors of parenting behavior. *Applied Developmental Science, 12*(4), 188–197. https://doi.org/10.1080/10888690802388110

Homem, T. M. F. C. (2014). *A Eficácia de um Programa de Intervenção Parental para Pré-escolares com Comportamentos de Oposição: O Programa Anos Incríveis* [Doctoral thesis, Coimbra University]. Repositório Científico da Universidade de Coimbra https://eg.uc.pt/handle/10316/26314.

IntHout, J., Ioannidis, J. P. A., Rovers, M. M., & Goeman, J. J. (2016). Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open, 6*(7), Article e010247. https://doi.org/10.1136/bmjopen-2015-010247

Jeong, K., Franchett, E. E., de Oliveira, C. V. R., Rehmani, K., & Yousafzai, A. K. (2021). Parenting interventions to promote early child development in the first three years of life: A global systematic review and meta-analysis. *PLoS Medicine, 18*(5), Article e1003602. https://doi.org/10.1371/journal.pmed.1003602

Karjalainen, P., Kiviruusu, O., Aronen, E. T., & Santalahti, P. (2019). Group-based parenting program to improve parenting and children's behavioral problems in families using special services: A randomized controlled trial in a real-life setting. *Children and Youth Services Review, 96*, 420–429. https://doi.org/10.1016/j.childyouth.2018.12.004

Larsson, B., Fossum, S., Clifford, G., Drugli, M. B., Handegård, B. H., & Mörch, W. T. (2009). Treatment of oppositional defiant and conduct problems in young Norwegian children: Results of a randomized controlled trial. *European Child and Adolescent Psychiatry, 18*(1), 42–52. https://doi.org/10.1007/s00787-008-0702-z

Leijten, P., Melendez-Torres, G. J., Knerr, W., & Gardner, F. (2016). Transported versus homegrown parenting interventions for reducing disruptive child behavior: A multilevel Meta-regression study. *Journal of the American Academy of Child & Adolescent Psychiatry, 55*(7), 610–617. https://doi.org/10.1016/j.jaac.2016.05.003

Leijten, P., Raaijmakers, M. A. J., de Castro, B. O., van den Ban, E., & Matthys, W. (2017). Effectiveness of the incredible years parenting program for families with socioeconomically disadvantaged and ethnic minority backgrounds. *Journal of Clinical Child and Adolescent Psychology, 46*(1), 59–73. https://doi.org/10.1080/15374416.2015.1038823

Leung, C., Sanders, M. R., Leung, S., Mak, R., & Lau, J. (2003). An outcome evaluation of the implementation of the triple P-positive parenting program in Hong Kong. *Family Process, 42*(4), 534–544. https://doi.org/10.1111/j.1545-5300.2003.00531.x

Lunkenheimer, E., Olson, S., Hollenstein, T., Sameroff, A., & Winter, C. (2011). Dyadic flexibility and positive affect in parent-child coregulation and the development of child behavior problems. *Development and Psychopathology, 23*(2), 577–591. https://doi.org/10.1017/s095457941100006x

Ma, L.-L., Wang, Y.-Y., Yang, Z.-H., Huang, D., Weng, H., & Zeng, X.-T. (2020). Methodological quality (risk of bias) assessment tools for primary and secondary medical studies: What are they and which is better? *Military Medical Research, 7*(1). https://doi.org/10.1186/s40779-020-00238-8

Martinez, I., Garcia, F., Veiga, F., Garcia, O. F., Rodrigues, Y., & Serra, E. (2020). Parenting styles, internalization of values and self-esteem: A cross-cultural study in Spain, Portugal and Brazil. *International Journal of Environmental Research and Public Health, 17*(7), 23–70. https://doi.org/10.3390/ijerph17072370

Maryanski, A. (2010). WEIRD societies may be more compatible with human nature. *Behavioral and Brain Sciences, 33*(2), 103–104. https://doi.org/10.1017/s0140525x10000191

Meija, A., Calam, R., & Sanders, M. R. (2015). A pilot randomized controlled trial of a brief parenting intervention in low-resource settings in Panama. *Prevention Science, 16*(6), 707–717. https://doi.org/10.1007/s11121-015-0551-1

Meija, A., Leijten, P., Lachman, J. M., & Parra-Cardona, J. R. (2017). Different strokes for different folks? Contrasting approaches to cultural adaptation of parenting interventions. *Prevention Science, 18*(6), 630–639. https://doi.org/10.1007/s11121-016-0671-2

Michelson, D., Davenport, C., Dretzke, J., Barlow, J., & Day, C. (2013). Do evidence-based interventions work when tested in the "real world"? A systematic review and Meta-analysis of parent management training for the treatment of child disruptive behavior. *Clinical Child and Family Psychology Review, 16*(1), 18–34. https://doi.org/10.1007/s10567-013-0128-0

Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ, 339b*(2535). https://doi.org/10.1136/bmj.b2535

Morelli, G., Quinn, N., Chaudhary, N., Vicedo, M., Rosabal-Coto, M., Keller, H., … Takada, A. (2018). Ethical challenges of parenting interventions in low- to middle-income countries. *Journal of Cross-Cultural Psychology, 49*(1), 5–24. https://doi.org/10.1177/0022022117746241

Niec, L. N. (2018). Parent-child interaction therapy: A transdiagnostic intervention to enhance family functioning. In L. Niec (Ed.), *Handbook of parent-child interaction therapy* (pp. 3–15). Springer. https://doi.org/10.1007/978-3-319-97698-3_1.

Ogden, T., & Hagen, K. A. (2008). Treatment effectiveness of parent management training in Norway: A randomized controlled trial of children with conduct problems. *Journal of Consulting and Clinical Psychology, 76*(4), 607–621. https://doi.org/10.1037/0022-006X.76.4.607

Page, M. J., Altman, D. G., Egger, M., Higgins, J. P. T., & Smith, G. D. (2022). Assessing the risk of bias in randomized trials. In I. M. Egger (Ed.), *Systematic reviews in health research: Meta-analysis in context* (3rd ed., pp. 55–73). John Wiley & Sons Ltd.. https://doi.org/10.1002/9781119099369

Patterson, G. R., Chamberlain, P., & Reid, J. B. (1982). A comparative evaluation of a parent-training program. *Behavior Therapy, 13*(5), 638–650. https://doi.org/10.1016/S0005-7894(82)80021-X

Pedersen, G. A., Smallegange, E., Coetzee, A., Hartog, K., Turner, J., Jordan, M. J. D., & Brown, F. L. (2019). A systematic review of the evidence for family and parenting interventions in low- and middle-income countries: Child and youth mental health outcomes. *Journal of Child and Family Studies, 28*, 2036–2055. https://doi.org/10.1007/s10826-019-01399-4

Rad, M. D., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences, 115*(45), 11401–11405. https://doi.org/10.1073/pnas.1721165115

Resnicow, K., Soler, R., Braithwaite, R. L., Ahluwalia, J. S., & Butler, J. (2000). Cultural sensitivity in substance use prevention. *Journal of Community Psychology, 28*(3), 271–290. https://doi.org/10.1002/(SICI)1520-6629(200005)28:3<271::AID-JCOP4>3.0.CO;2-I

Rojas, G., Martínez, V., Martínez, P., Franco, P., & Jiménez-Molina, A. (2019). Improving mental health care in developing countries through digital technologies: A mini narrative review of the chilean case. *Frontiers in Public Health, 7*, 391–400. https://doi.org/10.3389/fpubh.2019.00391

Sahithya, B. R., Manohari, S. M., & Vijaya, R. (2019). Parenting styles and its impact on children – A cross cultural review with a focus on India. *Mental Health, Religion and Culture, 22*(4), 357–383. https://doi.org/10.1080/13674676.2019.1594178

Sanders, M. R. (1999). Triple P-Positive Parenting Program: Towards an empirically validated multilevel parenting and family support strategy for the prevention of

behavior and emotional problems in children. *Clinical Child and Family Psychology Review, 2*(2), 71–90. https://doi.org/10.1023/A:1021843613840.

Sanders, M. R., Markie-Dadds, C., Tully, L. A., & Bor, W. (2000). The triple P-positive parenting program: A comparison of enhanced, standard, and self-directed behavioral family intervention for parents of children with early onset conduct problems. *Journal of Consulting and Clinical Psychology, 68*(4), 624–640. https://doi.org/10.1037/0022-006X.68.4.624

Scavenius, C., Chacko, A., Lindberg, M. R., Granski, M., Vardanian, M. M., Pontoppidan, M., … Eiberg, M. (2020). Parent management training Oregon model and family-based services as usual for behavioral problems in youth: A National Randomized Controlled Trial in Denmark. *Child Psychiatry and Human Development, 51*(5), 839–852. https://doi.org/10.1007/s10578-020-01028-y

Schuhmann, E. M., Foote, R. C., Eyberg, S. M., Boggs, S. R., & Algina, J. (1998). Efficacy of parent-child interaction therapy: Interim report of a randomized trial with short-term maintenance. *Journal of Clinical Child Psychology, 27*(1), 34–45. https://doi.org/10.1207/s15374424jccp2701_4

Sigmarsdóttir, M., Thorlacius, Ö., Guðmundsdóttir, E. V., & DeGarmo, D. S. (2014). Treatment effectiveness of PMTO for Children's behavior problems in Iceland: Child outcomes in a Nationwide randomized controlled trial. *Family Process, 54*(3), 498–517. https://doi.org/10.1111/famp.12109

Simons, R. L., Lin, K., Gordon, L. C., Brody, G. H., Murry, V., & Conger, R. D. (2004). Community differences in the association between parenting practices and child conduct problems. *Journal of Marriage and Family, 64*(2), 331–345. https://doi.org/10.1111/j.1741-3737.2002.00331.x

Smith, J. D., Cruden, G. H., Rojas, L. M., Ryzin, M. V., Fu, E., Davis, M. M., … Brown, C. H. (2020). Parenting interventions in pediatric primary care: A systematic review. *Pediatrics, 146*(1), Article e20193548. https://doi.org/10.1542/peds.2019-3548

Soto, A., Smith, T. B., Griner, D., Domenech-Rodríguez, M., & Bernal, G. (2018). Cultural adaptations and therapist multicultural competence: Two meta-analytic reviews. *Journal of Clinical Psychology, 74*(11), 1907–1923. https://doi.org/10.1002/jclp.22679

Stattin, H., Enebrink, P., Özdemir, M., & Giannotta, F. (2015). A national evaluation of parenting programs in Sweden: The short-term effects of using an RCT effectiveness design. *Journal of Consulting and Clinical Psychology, 83*(6), 1069–1084. https://doi.org/10.1037/a0039328

Sterne, J. A. C., Savović, J., Page, J. M., Elbers, R. G., Boutron, I., Cates, C. J., … Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ, 366*(l4898). https://doi.org/10.1136/bmj.l4898

Sundell, K., Beelmann, A., Hasson, H., & Schwarz, U. (2016). Novel programs, international adoptions, or contextual adaptations? Meta-analytical results from German and Swedish intervention research. *Journal of Clinical Child & Adolescent Psychology, 45*(6), 784–796. https://doi.org/10.1080/15374416.2015.1020540

Twisk, J., Bosman, L., Hoekstra, T., Rijnhart, J., Welten, M., & Heymans, M. (2018). Different ways to estimate treatment effects in randomised controlled trials. *Contemporary Clinical Trials Communications, 10*, 80–85. https://doi.org/10.1016/j.conctc.2018.03.008

United Nations Development Programme. (2020). Human development data center. http://hdr.undp.org/en/data.

United Nations Office on Drugs and Crime (UNOFC). (2009). *Guide to implementing family skills training programmes for drug abuse prevention.* New York, NY: United Nations.

Webster-Stratton, C. (2001). The incredible years: Parents, teachers, and children training series. *Residential Treatment for Children & Youth, 18*(3), 31–45. https://doi.org/10.1300/J007v18n03_04.

Webster-Stratton, C., Reid, M. J., & Hammond, M. (2004). Treating children with early-onset conduct problems: Intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child and Adolescent Psychology, 33*(1), 105–124. https://doi.org/10.1207/S15374424JCCP3301_11

Weeland, J., Chhangur, R. R., van der Giessen, D., Matthys, W., de Castro, B. O., & Overbeek, G. (2016). Intervention effectiveness of the incredible years: New insights into sociodemographic and intervention-based moderators. *Behavior Therapy, 48*(1), 1–18. https://doi.org/10.1016/j.beth.2016.08.002

Wiesz, J. R., Donenberg, G. R., Han, S. S., & Weiss, B. (1995). Bridging the gap between laboratory and clinic in child and adolescent psychotherapy. *Journal of Consulting and Clinical Psychology, 63*(5), 688–701. https://doi.org/10.1037//0022-006x.63.5.688

World Health Organization (WHO). (2010). *Violence prevention: The evidence.* Geneva, Switzerland: WHO.