

This is the accepted version (29 August 2018) of the following article: Nye, E., Melendez-Torres, G.J., & Gardner, F. Mixed methods systematic review on effectiveness and experiences of the Incredible Years Teacher Classroom Management program. *Review of Education*, X(X), XX-XX.

Mixed methods systematic review on effectiveness and experiences of the Incredible Years

Teacher Classroom Management program

Elizabeth Nye^{1*}, G.J. Melendez-Torres², and Frances Gardner¹

¹Centre for Evidence-Based Intervention, Department of Social Policy and Intervention,
University of Oxford

²Centre for the Development and Evaluation of Complex Interventions for Public Health
Improvement, School of Social Sciences, University of Cardiff

*Corresponding author: elizabeth.nye@spi.ox.ac.uk; Barnett House, 32 Wellington Square,
Oxford, OX1 2ER, United Kingdom; +44 1865 270325

ORCIDiDs:

Elizabeth Nye <https://orcid.org/0000-0001-8424-1984>

G.J. Melendez-Torres <https://orcid.org/0000-0002-9823-4790>

Frances Gardner <https://orcid.org/0000-0001-7508-7348>

Abstract

Children's early problematic behaviour correlates with later deviant behaviour. The Incredible Years Teacher Classroom Management (IY TCM) program trains teachers in proactive strategies to disrupt patterns of negative child behaviour. This mixed methods systematic review applies multilevel meta-analysis and qualitative meta-synthesis to answer: What is known about the effectiveness of IY TCM, and how do people experience the program and its effects?

We included RCTs comparing IY TCM against treatment-as-usual/waitlist controls and interview/focus group studies with stakeholders. We pooled effect sizes using multilevel meta-analysis and applied grounded theory analysis to qualitative data. We used framework analysis and integrative grids for cross-synthesis.

Nine studies from England, Ireland, Jamaica, the United States, and Wales were included. IY TCM significantly reduced teachers' use of negative classroom management strategies and reduced conduct problems among high-risk children. No significant effects were found for improving child prosocial behaviours. Qualitative findings illuminated iterative learning processes between the individual and group and identified a wider range of perceived benefits than was measured in trials. Cross-synthesis demonstrates both harmonious overlap and gaps from differing research priorities.

IY TCM presents an opportunity to lessen violence in schools, leading to improved learning environments for all. It reduced teachers' negative classroom management practices and high-risk children's aggressive/non-compliant behaviours. Experiential data mirrors these findings and expands upon them with a wide range of perceived benefits, demonstrating IY TCM

is acceptable and enjoyable. Future quantitative research should seek to test the processes produced in the qualitative meta-synthesis, to better understand mechanisms of change.

Keywords: classroom management; conduct problems; systematic review

**Mixed methods systematic review on effectiveness and experiences of the Incredible Years
Teacher Classroom Management program**

Children's mental health issues are a substantial public health issue globally. Conduct disorder, diagnosed as early as the preschool years, involves aggressive and noncompliant behaviours beyond the normal range expected from children and adolescents (American Psychiatric Association, 2013; World Health Organization, 1992). Children with diagnosed hyperkinetic (attention-deficit/hyperactivity, ADHD) disorder display heightened activity levels combined with difficulties controlling their behaviour. These behaviours are not limited to one environment (e.g., home, school) and are chronic rather than short-lived (American Psychiatric Association, 2013). Misbehaviour is not uncommon among children with ADHD, however this is due to "unthinking (rather than deliberately defiant) breaches of rules" (World Health Organization, 1992, p. 206), which is one way of differentiating ADHD from conduct disorder. However, conduct disorder often presents as comorbid with ADHD, and the combination of both of these disorders predicts poorer outcomes (American Psychiatric Association, 2013).

Although there are some variations across global context, prevalence rates indicate that upwards of 10% of school-age children suffer from clinical-level difficulties across conduct and hyperkinetic domains. A 2011 European review of epidemiological studies estimated prevalence rates of 3% (95% CI [0.0-6.0]) for conduct disorder and 5% (95% CI [1.0-17.8]) for hyperkinetic disorder for school-age children between 5- to 17-years-old (Wittchen et al., 2011). In the United States (US), a nationally representative survey of school-age children 8- to 15-years-old identified over 2% (95% CI [1.8-2.4]) with conduct disorder and almost 9% (95% CI [7.9-9.3]) with ADHD (Merikangas et al., 2010). Research from a review of Brazilian studies likewise finds over 4% (95% CI [2.1-7.9]) of children under age 18-years-old identified with either

conduct disorder or oppositional defiant disorder (J. Murray, Anselmi, Gallo, Fleitlich-Bilyk, & Bordin, 2013).

However, these prevalence rates are limited to only those children who meet diagnostic criteria for *disorders*, and they do not account for the potentially many children who have heightened *problems* in any of these domains at sub-clinical levels. The Brazilian review established that an additional 20 per cent 95% CI [15.9-26.9] of children could be classified as having conduct problems (J. Murray et al., 2013), a number which is similarly supported in other countries (Snyder, 2001). In all, then, it is possible that almost one-third of children experience difficulties across these behavioural health domains.

Teachers' difficulties with classroom management

Teachers regularly report feeling underprepared for the reality of managing a classroom, especially those that include children identified with special educational needs. According to Evertson and Weinstein (2013), classroom management is the set of “actions teachers take to create an environment that supports and facilitates both academic and social-emotional learning” (p. 4). It is a multifaceted construct that includes both preventative (e.g., routines) and reactive (e.g., consequences) strategies (Korpershoek, Harms, De Boer, Van Kuijk, & Doolaard, 2016). Korpershoek et al. (2016) identified four primary purposes for classroom management programs: 1) changing teachers' behaviours, 2) improving the teacher-child relationship, 3) changing children's behaviours, and 4) improving children's social-emotional development, although these categories “are not mutually exclusive” (p. 646).

The focus of this mixed methods systematic review is the Incredible Years Teacher Classroom Management (IY TCM) program, which emphasizes proactive teacher classroom management by focusing on teacher attention, encouragement, praise, building positive

relationships, the writing of behaviour plans, and social skills/problem-solving training (Webster-Stratton, 2001). This program is delivered by two trained facilitators for small groups of teachers over the course of six, monthly full-day sessions, following a manual. IY TCM incorporates a variety of strategies (e.g., video vignettes, role play, and group discussion) that are common to all IY programs, and for which there is much research to support their use. The program is based on Patterson's childhood aggression theory (Patterson, DeBaryshe, & Ramsey, 1990), Bowlby's attachment theory (Bowlby, 1973), and Bandura's social learning theory (Bandura, 1978). Originally developed in the US, IY TCM has been implemented internationally, particularly in European countries such as Portugal (Webster-Stratton, Gaspar, & Seabra-Santos, 2012), Norway (Fossum, Handegård, & Britt Drugli, 2017), and Jamaica (Baker-Henningham, Scott, Jones, & Walker, 2012; Baker-Henningham & Walker, 2018). However, compared to the IY parent training program, which is extensively researched and reviewed, IY TCM has not been as comprehensively evaluated or reviewed in isolation from other IY programs.

The overall effectiveness of a program like IY TCM and an understanding of how it is experienced by key stakeholders has implications for how society addresses behavioural health difficulties among school-age children. Moreover, understanding the core mechanisms and key contextual features of successful IY TCM implementation, supports insights into adaptations needed to respond to contextual features. Considering the commitment to investing in infrastructure to widen access to education at all levels and the systems in place to train teachers to become effective educators (United Nations General Assembly, 2015), implementing a teacher training program as both a professional development opportunity for educators and a

public health intervention for children has the great potential to be an effective use of resources and appropriate avenue by which to reach a large number of children.

Aims and research design

Mixed methods systematic reviews are able to address broad research questions, which involve empirical and theoretical components, producing findings that are directly relevant to stakeholders. This mixed methods systematic review seeks to do just this – to both test effectiveness and examine the theoretical understanding of how IY TCM operates and is experienced by people and to illuminate the many ways by which it can be judged effective. To this end, this mixed methods systematic review aims to illuminate the effectiveness and experiences of the IY TCM program. To accomplish this aim, this systematic review answers three overarching questions:

1. What are the effects of IY TCM on teacher and child outcomes?
2. How do people experience/perceive the program and its effects?
3. When considered together, what conclusions can we draw based on the effectiveness and experiential data?

This study is conducted according to a mixed methods convergent parallel design, which involves a phase in which qualitative and quantitative data are concurrently yet independently collected and analysed, followed by a phase which sees these two strands joined together and data merged for subsequent interpretation and discussion (Creswell & Plano Clark, 2011; Teddlie & Tashakkori, 2009). By expanding the current review's remit to include not only trial evidence but also qualitative evidence, this review aims to provide a more comprehensive understanding of the effectiveness of IY TCM and how it is received by users to inform key decision-makers at various levels of practice and policy.

Method

A protocol for this review was registered in PROSPERO database (ref. no. blinded).

Criteria for Considering Studies for this Review

Types of quantitative studies. Studies included in the quantitative strand of this review must have met the following criteria:

1. Population: early childhood and primary school teachers and children ages three- to eight-years-old
2. Intervention: IY TCM
3. Comparison: treatment-as-usual and/or waitlist control
4. Design: randomised controlled trials

Studies that did not meet the above criteria (e.g., delivered IY TCM in conjunction with additional treatment or with no control group) were excluded from this review. The purpose of isolating IY TCM from other programs was to determine which outcomes are attributable to the IY TCM program alone and not in part or in full due to other programs.

The primary quantitative outcomes for this review were based on the stated outcomes for the IY TCM program:

1. Teacher classroom management strategies
2. Child conduct problems (e.g., aggressive or noncompliant behaviours)

Secondary outcomes also followed the stated intent of the IY TCM program and included:

1. Teacher collaboration with parents
2. Child prosocial behaviours
3. Child academic readiness (e.g., reading skills)

A *post hoc* child behaviour difficulties analysis was added to include not only all measures of conduct problems but also measures of hyperactivity and peer problems, which were reported alongside or separate from measures of conduct problems across different studies. This is reported separately in the supplementary files.

Types of qualitative studies. Studies included in the qualitative strand of this review should have:

1. Data from qualitative studies or mixed methods studies reporting qualitative outcomes;
2. Clearly stated that participants' responses pertained to IY TCM or were part of a larger study on this specific program;
3. Reported IY TCM data separately where IY TCM was combined with any other additional program.

Studies were excluded if data reported only on non-IY classroom management or teacher training programs or if the qualitative data reported on either the IY parent or child programs.

The qualitative strand in this review explored social processes and sought to understand interactions of key stakeholders with the content and delivery of IY TCM, as well as their subsequent views of the strategies and implementation of the techniques in their practice.

Participants in the qualitative studies were 'key stakeholders', and could have included those directly involved with IY TCM (e.g., class teachers, group facilitators) or those more indirectly involved (e.g., headteachers, teaching assistants, parents, children).

Search Strategy. One systematic search to identify relevant studies for both strands in this review, updated as of July 2018, was conducted across 13 electronic databases and four related websites (Table 1). Reference lists of included studies were searched, and researchers

who published papers regarding IY TCM or related topics were emailed to ask about potential studies.

[Table 1 near here.]

To increase specificity without compromising the sensitivity of any electronic database searches, the search terms focused on identifying IY TCM, which is a branded program that operates according to a manual. Initial search terms were first piloted, and the search terms were pared down to the most parsimonious set without compromising sensitivity.

Studies identified from electronic searches and personal contacts were combined, and duplicates removed. Two independent reviewers screened the titles and abstracts for potential relevance to this project. Potentially relevant reports were retrieved in full-text and assessed for eligibility using the predetermined inclusion criteria.

Data extraction and risk of bias assessment of included quantitative studies. Two reviewers independently extracted quantitative data into a spreadsheet using a uniform system. In cases where data were unclear or unreported, attempts were made to contact the study authors for clarification. Where the intraclass correlation coefficient (ICC) was reported for the whole sample (e.g., all observed children), but not for other reported subgroups (e.g., high- or low-risk children), it was assumed that the ICC for these subgroups was the same as for the whole sample in that study. Where the ICC was not stated in a study, the mean ICC on that specific outcome was calculated from the other included studies and used.

All studies included in the quantitative strand were assessed using the Cochrane Risk of Bias tool (Higgins & Green, 2008). In situations where multiple reports (e.g., both a published journal article and PhD dissertation) were available for an individual study, information available across all reports was considered when making decisions pertaining to risk of bias. In situations

such as lack of participant blinding (not possible with this intervention) or insufficient details regarding any of the different domains, a standard rating of unclear risk was assigned (Higgins & Green, 2008).

Quantitative multilevel meta-analysis. It is common for individual studies to report data from multiple measures for outcomes (e.g., teacher-reported and parent-reported child behaviour). Subsequently, effect sizes cluster within studies, raising the issue of data dependency. Multilevel meta-analysis avoids overestimating effect sizes by modelling the nested nature of effect sizes from included studies (Van Den Noortgate, López-López, Marín-Martínez, Sánchez-Meca, & Van Den Noortgate, 2013; Van Den Noortgate & Onghena, 2003). Multilevel analysis is preferable to traditional meta-analysis techniques, such as selecting one effect size per study, because it approximates an overall effect size (“grand mean”) on an outcome using all available relevant data. Selecting only one effect size per study for each outcome yields less statistical power, and introduces researcher bias due to judgment decisions as to which effect size might be best (Polanin, 2013).

Where possible, raw data reported in each included study was extracted and adjustments were first made for within study participant clustering, before calculating Hedges’ g effect sizes and estimating a multilevel meta-analysis to calculate the Hedges’ g grand mean (β_{00}). If group means, standard deviations, and group sizes were not reported, effect sizes and standard errors were extracted and transformed into Hedges’ g for uniformity. Where subgroups within a study were reported separately and not together (e.g., by age), an average effect size was calculated. Hedges’ g was calculated instead of Cohen’s d because it adjusts for biases resulting from small sample sizes (Hedges & Olkin, 1985), which was a factor with included data in this review.

Statistical analyses were run in R Studio (Version 3.2.2) using the metafor package (Viechtbauer, 2010).

Data extraction and quality assessment of included qualitative studies. Qualitative data were extracted into NVivo 11 for analysis. Qualitative studies were assessed for quality using an adapted EPPI-Centre checklist, which has been used in other qualitative meta-syntheses (Bonell et al., 2013). This checklist includes domains on rigor in sampling, data collection, data analysis, whether findings were grounded in the data, and the breadth/depth of the findings. It also considers reliability, trustworthiness, and usefulness of the study to the review. Studies were not excluded for poor quality; however, ‘low quality’ studies did not serve as the initial point for reciprocal translation in the meta-synthesis process. Where findings from qualitative data were sparsely reported, further elaboration from authors was not sought, because the qualitative meta-synthesis is based on the authors’ interpretations/findings, and to ask for specific elaboration on specific areas could have changed the perspectives of these findings.

Qualitative meta-synthesis. Analysis for this qualitative meta-synthesis was based in a grounded theory approach, which operated both inductively and deductively and aimed to generate, not test, theory (Glaser & Strauss, 1967). When coding qualitative data, analysts transformed data from its original state to a conceptual level (Corbin & Strauss, 2008). To start, open codes were assigned to the data, with the goal of identifying the many various inquiries one could pursue with the data (J. Green & Thorogood, 2014). Then, analysts connected categories by establishing their relationships to one another through axial coding (J. Green & Thorogood, 2014). The analysis proceeded towards progressively more abstract and theoretically informed codes with the emergence of core categories that remain substantiated in the original data (J. Green & Thorogood, 2014). Throughout this process, analysts continuously made comparisons

within and between incidents from the data. Sensitivity analyses were not applied to the qualitative strand, as specific methods for doing so are as yet unclear and the topic itself is contentious within the discourse about qualitative meta-synthesis (Nye, Melendez-Torres, & Bonell, 2016).

Cross-synthesis. After the quantitative multilevel meta-analysis and qualitative meta-synthesis were complete, the results were cross-synthesised to examine common and divergent themes and develop new implications for research and practice. Adapting the integrative grid from Flemming (2010), data for the cross-synthesis were charted in two ways. First, key results of the quantitative synthesis were inserted into column headings and the qualitative studies were inserted into the rows. Findings from the qualitative studies populated the cells, indicating the intersection between the quantitative outcomes and qualitative studies. Second, key themes from the qualitative synthesis were inserted into the column headings and the quantitative studies were put in each row. Results from the multilevel meta-analysis and narrative synthesis occupied the cells to illustrate the joining of qualitative findings and quantitative studies. By creating these integrative grids both ways, neither the effectiveness data nor the experiential data were valued over the other as the ‘starting point’ and discussion could subsequently emerge from both explanatory and exploratory positions. Visually, these grids also provided an illustration of which themes/concepts were commonly addressed between the quantitative and qualitative research (regardless of agreement or disagreement), which themes/concepts were only addressed by one of the two strands of research, and which studies covered a broad or narrow range of outcomes/topics.

Results

After conducting electronic database and website searches and contacting relevant researchers in this field, all retrieved records were imported into EndNote X7 and duplicates removed (Figure 1). Two researchers independently screened all titles and abstracts for the 6,075 remaining records. Based on this initial screening, 6,019 records were deemed irrelevant to the review, leaving a potential 56 records to be assessed for inclusion based on their full text. Two researchers independently read 56 full-text reports and judged each record to be included or excluded based on pre-specified criteria. Thirty-four records were excluded for not meeting inclusion criteria.

[Figure 1 near here.]

Nine studies reported in 22 records met inclusion criteria for this mixed methods review: seven studies in 15 records were classified as quantitative, and six studies in nine records were classified as qualitative. Only two records met inclusion criteria for both quantitative and qualitative strands of this review (Leckey et al., 2016; McGilloway et al., 2010). Studies in the quantitative strand took place in Jamaica (Baker-Henningham et al., 2012; Baker-Henningham & Walker, 2018), Limerick, Ireland (Hickey et al., 2015; Leckey et al., 2016; McGilloway et al., 2010), Missouri, US (Herman & Reinke, 2017; Reinke, Herman, & Dong, 2014, 2016), North Carolina, US (D. Murray, Murr, & Rabiner, 2012; D. Murray, Rabiner, & Carrig, 2014; D. W. Murray, Rabiner, Kuhn, Pan, & Sabet, 2018), Rhode Island, US (Shepard, Doctoroff, Dickstein, & Seifer, 2008), and Wales (Hutchings, Martin-Forbes, Daley, & Williams, 2013; Martin, 2009). Studies in the qualitative strand took place in Dublin, Ireland (Davenport & Tansey, 2009; Kennedy, 2016), England (Marlow et al., 2015), Jamaica (Baker-Henningham & Walker, 2009),

Limerick, Ireland (Hyland, 2014; Leckey et al., 2016; McGilloway et al., 2010, 2012), and Wales (Hutchings et al., 2007).

Each of the nine studies is referred to by its location in the results because three studies contributed evidence to both quantitative and qualitative strands (Jamaica, Limerick, Wales) and all but three studies were reported across multiple records. All but one of the included studies were identified through the database/website searching; only the quantitative report from England was identified by contacting researchers in the field due to its status as in press when the electronic search was conducted (although two protocols for this study were identified through the electronic search).

Characteristics of Included Studies

The seven included studies in the quantitative strand together describe the results of testing IY TCM in more than 147 schools, over 336 teachers and teaching assistants and with 5,759 children (Table 2). The schools involved in these studies were located across rural and urban areas. In four studies (England, Limerick, Jamaica, and Wales) baseline data were collected prior to randomisation, allowing researchers to identify a subset of children from each classroom based on their level of risk for behavioural problems. The Missouri and North Carolina studies both block randomised in clusters. In the Rhode Island study, researchers blind to condition collected baseline data following randomisation. Three studies reported specific exclusion criteria for children beyond lack of parental consent. The England study required that the children and their parents have “sufficient English language comprehension to understand recruitment information and complete outcome measures” (Ford et al., 2018, p. 3). The England study also excluded schools that were predominantly for children with special educational needs (SEN), did not have stable school leadership via a headteacher or were rated as ‘failing’ from

their last school inspection. Children with SEN who were in mainstream English schools were included in the England study, though. The Limerick study excluded children who were absent from school more than once during data collection and those identified as having a developmental disorder. The Jamaica study excluded children who were living in an institution or who had a sibling enrolled in the study, those children with attendance less than 70 per cent, and any with identified developmental disorders.

[Table 2 near here.]

All included studies provided IY TCM using trained facilitators. In all studies, facilitators had at minimum been trained in IY TCM processes and had backgrounds in related fields, including psychology, education, and school social work. In the US studies, the facilitators had been trained by the program developer. In the England study, facilitators had delivered at minimum two previous TCM groups prior to the study, and they received monthly supervision sessions with the program developer. In the Limerick study, the primary facilitator was a certified group leader, indicating that they had been accredited as having expertly delivered the IY TCM program and completed the certification process. In two of these studies (North Carolina and Wales) the primary facilitator was a certified IY mentor, meaning that not only were they accredited as having expertly delivered the program, but additionally, they were authorized to provide IY training and supervise group leaders whom they had trained. The England, Missouri, and Rhode Island studies delivered IY TCM to groups of teachers over the course of six sessions, while the Limerick, North Carolina, and Wales studies delivered it in five sessions. In between the monthly training sessions, group facilitators in the Limerick study called the intervention teachers once to discuss implementation of learned strategies, and participating teachers in the North Carolina study were able to meet with their group facilitators

for individual consultation on implementing the learned strategies in their classrooms. The Jamaica study culturally adapted IY TCM based on qualitative focus groups, a pilot study, and in-depth qualitative interviews post-pilot study (Baker-Henningham, 2011; Baker-Henningham, Walker, Powell, & Gardner, 2009a, 2009b), with training occurring over eight sessions, and with practical activities tailored to reflect a Jamaican classroom rather than the original American classrooms. In addition to the training, teachers received one hour of in-class coaching each month for four months.

Most studies treatment-as-usual wait-list controls, where teachers attended regularly scheduled trainings during the trial, e.g. Ministry of Education or Head Start practices. The Limerick and Wales studies had wait-list controls, with the Wales study specifying that control teachers should not have in-service or other professional development training sessions offered during the trial.

The six included qualitative studies focused predominantly on gaining insights into experiences of 111 participating IY TCM teachers, exploring teacher perceptions of processes (i.e., barriers and facilitators to implementation) and looking at recommendations for the future (Table 4). Two studies interviewed other stakeholders beyond teachers, focusing on the experiences of IY TCM group facilitators (Limerick) or the perceptions of school headteachers (England). Sample sizes varied between included studies and ranged from around 12 (Dublin 2) to 47 (England) depending on the number of participating teachers, group facilitators, and headteachers. Typically, researchers conducted individual interviews (Limerick, Jamaica, Wales), focus groups (England), or both (Dublin 2) with the teachers, although the Dublin 1 study had teachers complete qualitative questionnaires. Group facilitators participated in individual interviews (Limerick), and head teachers participated in semi-structured telephone

interviews (England). All participants provided their experiences and views immediately following completion of IY TCM, and the Limerick study followed up with teachers six months post-intervention to gain additional insights about how perceptions endured. Interview and focus group data were audio-recorded and transcribed for analysis in the included studies. Researchers applied framework analysis (England, Limerick, Jamaica), thematic content analysis (Dublin 2, Wales), and frequency counts (Dublin 1) to their data.

[Table 4 near here.]

Excluded Studies

Thirty-four records were examined and excluded from this review for failing to meet all of the *a priori* inclusion criteria set forth for either the quantitative or qualitative strands (Supplementary File 1). Most quantitative records were excluded due to study design (i.e., not RCTs) or intervention (i.e., not IY TCM). Studies commonly evaluated either IY TCM in a pre-/post- no control design (Carlson, Tiret, Bender, & Benson, 2011; Fergusson, Horwood, & Stanley, 2013) or combined IY TCM with other interventions (Baker-Henningham et al., 2009a; Fossum et al., 2017), limiting conclusions that could be drawn regarding to what extent post-test results were due specifically to the implementation of IY TCM and would not have occurred either without intervention or due to other programs. Most qualitative records were excluded because they were reviews of the literature and not primary qualitative studies (Hutchings, 2012; Pidano & Allen, 2015).

Findings

Experiences of key stakeholders. Analysing findings from each included qualitative study in this qualitative strand produced an overarching understanding of how key stakeholders (predominantly classroom teachers) experienced IY TCM (Figure 5). Throughout the entire

course, teachers experienced primary learning opportunities (first-hand practice) and secondary learning opportunities (learning through others). Ultimately, this led to their perceptions of the effects of IY TCM. Outcomes described by key stakeholders are explained as they pertain to teachers, classrooms, and children, reflecting the hierarchical relationship existing in educational environments, whereby children are nested within classrooms with teachers (for illustrative quotations, see Table 5). Outcomes also include experienced critiques of the IY TCM course.

[Figure 5 near here.]

[Table 5 near here.]

Cycle of learning. The cycle of learning emerged not only from the design of IY TCM but also in processes experienced by teachers throughout the academic year. The intersection of individual and group was evident from the beginning, and IY TCM group facilitators were crucial in ensuring smooth group integration. Even when teachers felt that the delivery did not always match their needs, they remained positive about the facilitators. Each teacher arrived with previous knowledge and skills, and initial group relationships needed to be established and a positive group dynamic fostered. The value of shared learning experiences emerged not only from joint problem-solving, but also in the emotional journey of moving from feeling isolated in a perceived weakness (i.e., poor classroom behaviour management) to the realisation that other teachers also experience these emotions (i.e., normalisation of the job challenges).

Over time, teachers communicated that their teaching practice was realigning with principles that they had originally learned, and which were being re-taught on the IY TCM course. Teachers felt they were able to go back to the basics and re-establish positive routines for classroom management that had slipped over time. The content and materials were also important, in that they could be compiled into a resource bank of classroom management

strategies for teachers to use in the future. This was crucial, because, teachers receive a new class of pupils with each academic year, and the types of challenging behaviours they might have present in their classroom inevitably differ.

Therefore, the cycle of learning involved not only the interchange between individual teachers and the training group as a whole, but also an extended feedback loop of direct learning, practice in the group, practice in the classroom, individual reflection, group sharing and reflection, re-setting one's approach to classroom behaviour management, establishing a support network, and building a resource bank of strategies for current and future use. Additionally, teachers reported experiencing positive effects from IY TCM training and feeling that it was worthwhile. Observing children's compliance and cooperation after implementing new strategies served to positively reinforce IY TCM strategies, and their experiences of IY TCM outcomes formed the second part of this overarching model.

Intervention Outcomes – Measured and Perceived. In addition to their experiences *learning* from IY TCM, the teachers also provided insights into their experiences of the *benefits* of training. Many intervention outcomes were also quantitatively measured. All of the included quantitative studies reported outcomes immediately post-treatment and, except for the Rhode Island study, reported adjusting their analyses to account for the clustered nature of their data. Data from the Rhode Island study, reported only in conference poster form, were incomplete and unable to be included in the multilevel meta-analysis. All teacher and student data were continuous. Results of the multilevel meta-analysis are reported in Table 3 and described by outcome below, alongside the perceived intervention outcomes from the qualitative strand.

[Table 3 near here.]

Teacher outcomes. Four studies reported data on teachers' strategies (Jamaica, Limerick, North Carolina, Wales) using the Teacher-Pupil Observation Tool (T-POT) (Martin, 2009; Martin et al., 2010), the Teacher Strategies Questionnaire (TSQ) (Webster-Stratton, 2005), or the Teacher Coder Impressions Inventory (TCI) (Webster-Stratton, Reid, & Hammond, 2001). There was a moderate, statistically significant effect of the IY TCM intervention on reducing teacher negative classroom management strategies from eight effect sizes across the Jamaica, Limerick, and Wales studies ($g = -0.49, [-0.89, -0.09]$), with moderate heterogeneity at the level of the effect sizes but not at the study level. Along with the North Carolina study, these studies also contributed 11 effect sizes on positive classroom management strategies. There was a very large, but non-statistically significant effect of the IY TCM intervention on increasing teacher positive classroom management strategies ($g = -0.73, [-1.92, 0.46]$) (Figure 2). For this outcome, moderate heterogeneity is present at both effect size and study levels, which likely contributes to the large confidence interval around the point estimate. Teacher collaboration with parents could not be meta-analysed.

[Figure 2 near here.]

The perceived outcomes experienced by the teachers fell into four related themes: (1) knowledge, (2) locus of control, (3) emotional well-being, and (4) teaching philosophy and practice (Figure 6). Broadly speaking, teachers experienced an increase in professional knowledge as a result of the course, including information about children's abilities/needs and a range of classroom management strategies. Equipped with this knowledge and the theory behind the positive strategies, teachers increasingly experienced greater self-confidence and a wider locus-of-control over their ability to deal with day-to-day challenging behaviours. This control led teachers to feel calmer and less stressed in the classroom and subsequently to experience

improved emotional well-being in their jobs. Over time and throughout the learning process previously discussed, teachers modified their teaching philosophies to emphasise positive strategies.

[Figure 6 near here.]

The teachers particularly noted the benefit of information about children's developmental abilities and needs in understanding why children acted in challenging ways and information on how teachers could best respond to children's actions, instead of reacting in the moment. By discussing the theories behind this knowledge, teachers felt more secure in this information, perhaps because they felt their understanding went deeper. Indeed, data from Limerick's long-term follow-up indicates that increased knowledge persisted into the next academic year.

Teachers also reported that their emotional well-being also benefited from participation in IY TCM. They were able to find more enjoyment in their interactions with children and described this as due to being able to better regulate their own emotions to remain calmer during fraught situations. As a result, they experienced less stress and isolation while on the job. These findings at post-intervention were also seen at six-month follow-up in the Limerick study.

Ultimately, the collection of teacher outcomes in knowledge, locus-of-control, and emotional well-being can be seen in their refreshed teaching philosophy and practice. Teachers reported experiencing a new and improved passion for teaching and reliance on positive strategies. Teachers expressed that they used to feel this way about teaching and were returning their daily practice back to their ideal approach to teaching. However, being on the IY TCM course and observing positive changes in their classrooms led teachers to reset their professional attitude and focus. As with the other teacher outcomes, the change in teaching philosophy continued through long-term follow-up.

Classroom outcomes. Perceived classroom outcomes refer to qualitative data on whole-group outcomes of both teachers and children. Two themes emerged in this area: (1) relationships and (2) classroom environment. Reports of improved teacher-child and child-child relationships, at the classroom level as opposed to an individual (target child) level reflect the move by teachers to focus on positives exhibited in the classroom rather than negatives. With a broadening sphere of encouragement and co-operation at the classroom level, teachers also experienced improved relationships with the parents of their pupils.

Child outcomes. Four studies (England, Limerick, Missouri, Wales) reported data on all children (with baseline rates of behaviour difficulties ranging from low- to high-risk) using the T-POT and the Strengths and Difficulties Questionnaire (SDQ) (Goodman, 1997). Analysis of 12 effect sizes resulted in a small effect without statistical evidence of benefit ($g = -0.05$, $[-0.13, 0.03]$) and negligible heterogeneity at both effect size and study levels.

Analysis of only the high-risk children used data from three studies (Jamaica, Limerick, Wales) is based on the SDQ, T-POT, Dyadic Parent-Child Interaction Coding System (DPICS) (Eyberg & Robinson, 1981) and Multi-Option Observation System for Experimental Studies (MOOSSES) (Tapp, Wehby, & Ellis, 1995) measures. There was a moderate, statistically significant effect of the IY TCM intervention on reducing high-risk child conduct problems ($g = -0.35$, $[-0.49, -0.21]$), again with no heterogeneity detected at either level (Figure 3).

[Figure 3 near here.]

Teachers qualitatively reported experiencing reductions in children's challenging behaviours such as inattention, disruption, and aggression, and these outcomes were directly attributed to using classroom management strategies learned through IY TCM. Headteachers also noticed receiving fewer child referrals for discipline. Teachers felt that children were able

to stay focused and were generally more appropriately behaved while in the classroom.

According to the qualitative findings, teachers grouped aggression/noncompliance (i.e., conduct problems) together with attention/hyperactivity into the broader category of ‘behaviour difficulties’.

As a result of the teachers reporting a general perceived improvement in ‘behaviour difficulties’ and because not all effect sizes for child conduct problems isolated conduct problems from other constructs, such as hyperactivity and peer problems, an exploratory analysis (*post hoc*) examined the effect of IY TCM on child behaviour difficulties, operationalised as conduct problems, hyperactivity, and peer problems. Results from these analyses are available in Supplementary File 1, with statistically significant moderate effects for high-risk children ($g = -0.37$, [-0.46, -0.27]) and small but significant effects for all children ($g = -0.11$, [-0.20, -0.02]). Both of these analyses had little to no heterogeneity at effect size and study levels.

Analysis of eight effect sizes from five studies (England, Limerick, Missouri, North Carolina, Wales) included data using the T-POT, SDQ, and the Revised Teacher Social Competence scale (R-TSC) (Conduct Problems Prevention Research Group, 1995) measures. There was a small, non-statistically significant effect of the IY TCM intervention on increasing child prosocial behaviours ($g = -0.12$, [-0.30, 0.06]). Heterogeneity was not present at the level of effect sizes and small at the study level.

Analysis of four effects from only high-risk children in three studies (Jamaica, Limerick, Wales), was based on the SDQ and T-POT measures. The resulting moderate effect did not reach statistical significance in increasing high-risk child prosocial behaviours ($g = 0.31$, [-0.65, 0.03]), with moderate heterogeneity at the level of effect sizes (Figure 4).

[Figure 4 near here.]

Teachers also qualitatively reported feeling that children's social skills benefited from new strategies teachers had learned through IY TCM. Teachers reported experiencing that children co-operated better with each other and were able to problem-solve among themselves when conflict arose, rather than require the teacher's intervention to quell arguments. These perceived outcomes were seen not only among children who were considered to be lacking in such skills at baseline, but also by children who would not have been identified by the teacher as problematic. This particular finding is notable because it suggests that all children can benefit from these positive management strategies.

Not only had children's social skills grown, but teachers qualitatively reported feeling that children's emotional well-being was boosted as a result of the training the teachers had received. Teachers perceived children as demonstrating greater abilities in recognizing how they were feeling, naming the emotions, and responding appropriately in various emotional states.

Data on children's academic readiness and educational attainment were only reported in the two studies (Missouri, North Carolina), with anticipated reports in the future from the England study. This outcome was not meta-analysed, and both studies found no statistically significant effects of IY TCM intervention on children's academic areas. Qualitatively, the Dublin 2 and Jamaica studies touched on teachers' perceptions of children's outcomes in academics, where teachers reported perceiving an increased engagement in school among children, feeling that children's academic skills had improved as a result of the implementation of IY TCM strategies.

Critiques of the program. Despite the many positive experiences shared by the various stakeholders, critiques of IY TCM were also present and predominantly appeared in qualitative sections about recommendations for future implementation, barriers to implementation, and

difficulties with the program. Unlike previous meta-synthesis themes, these critiques were less cohesive across studies; it appears that there was greater variability within and across studies on whether participants experienced certain aspects of the program as negative.

Despite this relative dearth in data, a preliminary model of critiques is presented here, using a cyclical process of recruitment, delivery/content, outcomes, and scale-up to illustrate the difficulty of balancing competing aspects of program delivery (for figure, see Supplementary File 1). This model situates critiques within the implementation process so that decision-makers can ascertain how these might resonate in their own context and consider appropriate responses to their specific situation.

Reason for taking the course. Motivation for participating in IY TCM was highlighted as an important factor, with the distinction between teacher choice to participate contrasted against compulsion (e.g., headteacher assigns teacher to training). Group facilitators (Limerick) felt that teachers who willingly participated in IY TCM training would be more open to it compared to those who were compelled to participate. This was reported as a facilitator in the Dublin 2 study, where teachers “considered that their interest in upskilling as well as their openness and willingness to reflect on and change their practice were key enabling factors” (Kennedy, 2016, p. 105). The England study observed a variety of ways by which teachers were recruited to participate in IY TCM, although researchers did not observe cause for concern that teacher choice (or lack thereof) resulted in different experiences. This suggests that voluntary participation might be a moderator for teacher engagement and outcomes, but would need to be tested. The influence of teacher choice in participating on the training program and any potential change in effectiveness has very real implications for how decision-makers move forward when scaling up implementation of IY TCM at different levels.

IY TCM process and content. Among teachers who participated in IY TCM, there were three main critiques of process: its manualised nature, specific learning techniques, and the presence of extra challenges. Notably, the debate around fidelity and adaptation is present among qualitative data. Group facilitators expressed strong views in favour of implementing IY TCM in accordance to the manual, because they felt that implementation fidelity was important. Teacher feelings were less uniform: some appreciated the ‘step-by-step’ approach and others expressed dissatisfaction with what they felt was an inflexible curriculum that did not adequately adapt to meet their levels of experience. There were also some concerns over repetition of concepts and unfavourably long training days. It is unclear whether factors such as facilitator skills, teachers’ level of experience, or current classroom needs meaningfully contributed to differences in teachers opinions about the scope and sequence of the course. Moreover, if flexibility were introduced, questions arise regarding how to avoid omitting or reducing the dose of core components.

When it came to specific learning techniques used in the process of IY TCM, two techniques stood out as eliciting very strong opinions among participating teachers: role playing and video vignettes. Only some teachers were reported as expressing reluctance to engage in role play as a form of practicing strategies, but this learning strategy is worth mentioning due to the strength of the reluctance among those who raised it as an issue. Although an aversion to role play could perhaps stem in part from individuals’ shyness or embarrassment, in at least one case a teacher reported feeling patronized by the activity. The discomfort felt by participants when faced with role playing tasks highlights the importance of an earlier theme from the cycle of learning: establishing the group dynamic and creating a safe space. Included studies used facilitators who were at least basically trained to deliver IY TCM, and in some cases expertly

trained not only to deliver the program but also to train other facilitators. Although the included studies do not link facilitator expertise with teacher-reported responses regarding strategies such as role play, this could be one explanation. Indeed, despite strong reservations towards role play, teachers also identified this technique as a valuable part of the learning experience, suggesting that the most challenging aspects of learning could be particularly fruitful.

Unlike critiques of role playing techniques in the learning process, criticisms of video vignettes centred primarily on cultural differences. In all five studies where video vignettes were criticised, participants noted that the American classrooms portrayed in the vignettes were incongruent with their own settings. Despite acknowledging that video vignettes still served as conversation starters for IY TCM topics, teachers raised important questions regarding acceptance of and engagement with the program through appropriate video vignettes.

The Jamaica study reported on challenges in dealing with teachers' misinterpretation how/when to apply IY TCM strategies in the classroom as well as disagreeing with core tenets of IY TCM. Although other studies referenced the importance of time to rehearse and imbed strategies into teachers' day-to-day practice, none raised concerns about serious confusion or fundamentally incompatible approaches to classroom management, as were found in the Jamaican context. It is of particular note that in the Jamaica study, teachers were qualified at the paraprofessional (i.e., teaching assistant) level, which is in contrast to the further preparation and certification that would have been present among teachers in all of the other included study contexts (Ireland, England, and Wales). This discrepancy in prior professional training could explain in part why some Jamaican teachers struggled to appropriately apply strategies such as ignore, over and above needing time to practice. Furthermore, the authors note that 'Different cultural beliefs and values may also affect the acceptability of the goals and/or strategies

advocated by these programs' (Baker-Henningham & Walker, 2009, p. 633). This is particularly relevant for the disagreement over the use of corporal punishment as a strategy used by teachers in their classrooms. However, it should be noted that quantitatively, teachers from the Jamaica study experienced some of the strongest effects in both reduction of negative strategies and increase in positive strategies, possibly suggesting that contexts where teachers have lower baseline levels of training present opportunities for greatest growth.

Child outcomes. Despite the optimistic and encouraging litany of teacher-reported child benefits, teachers were concerned over a lack of attention on how to implement IY TCM strategies with older children compared to younger ones. Also, teachers questioned the effectiveness of IY TCM, suggesting it might be limited for identified subgroups of children whose needs were so great or different that they would not benefit from the application of these strategies alone.

IY TCM was originally designed to address children with clinical and pre-clinical problematic behaviours in school, suggesting that it is meant for high-risk children. Programs demonstrating effectiveness at the sample-level are assessed by difference in means, and do not represent individual trajectories or even suggest that all children will benefit. If subgroups of children are systematically not benefiting from application of IY TCM strategies, it could suggest that they require additional or different supports. The evidence available in this synthesis is insufficient to draw any conclusions regarding whether subgroups of children (e.g., those with special educational needs, in upper primary school) are indeed systematically underserved with IY TCM techniques. Nevertheless, there is preliminary experiential evidence that effects of IY TCM might be moderated by children with complex or special educational needs and those who are older.

Wider support and scaling-up. Finally, stakeholders expressed their views on the role of wider support for training and scaling-up implementation beyond each study's limited remit. This theme is incorporated under the critiques of the program synthesis because it illustrates the constraints present in the real-world, which could moderate program effectiveness. Additionally, it demonstrates a need to consider the level of implementation and contributes to the theory that intervention might need to occur at an additional, higher level of cluster: the school rather than the classroom. Group facilitators reported that the heightened interest of school headteachers was at least in part responsible for teacher motivation. Other school-based interventions, such as the PATHS curriculum, which target children's mental health have found that programs implemented with high fidelity are differentially effective due to the strength of head teacher (principal) support. That is, when good program implementation is coupled with head teacher support, greater reductions in conduct problems and increases in emotional skills were found (Kam, Greenberg, & Walls, 2003).

Despite these studies taking place across different countries and cultures, which produced contradictory critiques and recommendations in previous themes, the consistent message of widening the training to include all teachers and staff suggests that the theory of change might need to be expanded to the school level, rather than classroom level. According to some stakeholders, inconsistencies in other staff's management strategies hindered perceived effectiveness of IY TCM on child outcomes. This highlights the importance that teachers placed on the school as a whole community, rather than just their classroom as isolated sub-communities within schools. If the involvement of school leaders and implementation of whole-school IY TCM is a better approach to intervention, then it is entirely possible that results from

effectiveness studies may be underestimated because the program has not been comprehensively rolled-out.

However, wide-scale implementation of any training brings the conversation back around to the original critiques of the program: reason for taking the course. As already presented, those teachers who had a choice to participate on the training course quite possibly were more engaged in the process and open-minded to the content compared to those teachers who were obliged to attend by their school head. When the suggestion arises, as it has, that all teachers and school staff should have access to this course because it was so well received by study participants, this raises questions of how to do so without making it compulsory and risking engagement.

Results of the cross-synthesis

Concepts that were commonly covered across both quantitative and qualitative strands produced harmonious findings; data from each strand suggested similar trends or results. Generally, the quantitative evidence supported conclusions that IY TCM intervention is more effective than standard practice across studies, and qualitative teacher reports from these studies confirmed that teachers feel their practice and children's behaviours have improved.

The novel findings from this cross-synthesis lie predominantly in the areas that are not covered (or only partially addressed) by both strands. This dissonance illuminates different research priorities and ensuing gaps in our understanding of the processes and results of implementing the IY TCM program. Notably, the quantitative evidence lacks a focus on revealing and evaluating the possible mechanisms of change by which the intervention achieves its intended outcomes. Meanwhile, the qualitative evidence is missing input from most of the indirect key stakeholders (e.g., parents, children), relying almost entirely on views from teachers who have successfully completed (and not dropped out of) the training course. Furthermore,

while in some cases the qualitative research was conducted as part of feasibility or piloting work to inform a larger quantitative trial, the dearth of quantitative data on many of the themes that emerged from the qualitative data suggests that the integration of evidence across the quantitative-qualitative divide has been limited, and future work should seek to more fully and continuously weave together knowledge and experience gathered in both numerical and non-numerical forms. For the full integrative grids, see Supplementary File 2.

Teacher outcomes. The qualitative strand revealed four key areas in which teachers felt they had benefitted from the IY TCM program: 1) knowledge, 2) locus-of-control, 3) emotional well-being, and 4) teaching philosophy and practice. Only the fourth area received meaningful coverage from quantitative studies, with all studies reporting on teaching practice. Thus, contrary to the dearth of evidence across those first three areas, there is a strong agreement between qualitative and quantitative findings on the benefits of IY TCM on teachers' classroom practice. However, the process traced through the various teacher outcomes in the qualitative strand highlighted the fact that the other three identified teacher outcomes were considered valuable by participants. With little or no quantitative evidence available to ascertain the effectiveness of IY TCM on teachers' knowledge, locus-of-control, or emotional well-being, we cannot draw conclusions about any secondary effects this intervention might have. Even if schools might choose to train their teachers in IY TCM to change teacher classroom management practices, knowing about possible secondary outcomes could serve as additional motivation for decision-makers to invest in this type of professional development for their educators. Furthermore, the qualitative strand explains a possible theory-of-change for teacher practice, whereby teachers' practice is mediated by gains in knowledge, widening of their locus of control, and/or feeling emotionally calmer and less stressed. These are hypotheses that could be tested in

future studies, although the confidence that we can place in these hypotheses is lessened for not having any preliminary quantitative evidence to indicate whether these areas of teacher development quantitatively change post-IY TCM.

All five qualitative studies reported findings related to teachers' improved skills in positive classroom management strategies. These qualitative reports converge with the results from the multilevel meta-analysis of a moderate effect of IY TCM on increasing positive strategies. In contrast, only two qualitative studies (Dublin 2, Jamaica) reported teachers' experiences reflecting on their use of negative classroom management strategies, even though quantitative results showed a moderate, statistically significant effect of IY TCM on decreasing negative strategies. These data raise questions about why there is such a difference in attention to positive/negative strategies in the qualitative data. Were teachers more aware of their use of positive strategies than their reduction of negative strategies in day-to-day practice? Did researchers inadvertently phrase questions in a way that elicited responses on positive rather than negative strategies? Or might researchers have attached greater explanatory power to data on positive strategies than negative ones during analysis? This is unclear. Yet, the Jamaica qualitative study might differ from the others in its focus on decreased negative strategies due to contextual factors whereby corporal punishment was an acceptable and commonly used teacher strategy at baseline.

Child outcomes. Five key areas emerged in the qualitative strand regarding benefits of IY TCM for children: 1) behaviour, 2) social skills, 3) emotional well-being, 4) academics, and 5) other environments. The first three of these areas were most robustly supported and similar coverage trends are seen among quantitative studies. There is strong agreement within quantitative studies to support the qualitative constructs of improved child behaviour and social

skills. All studies examined these two constructs in some way, indicating the centrality of child behaviour and social skills as desired and valued outcomes from both the teachers' perspectives as well as the researchers' agendas. As further results from studies are published, effects on children's emotional well-being and academics should also be meta-analysed.

Critiques. Through the qualitative strand findings, several critiques of IY TCM were raised by participants (mainly teachers, but also group facilitators). Although these critiques were less consistent across participants and between studies, they introduced important issues for discussion. The critiques around the process and content of IY TCM are all situated within the fidelity-adaptation debate. Specifically, how flexible can group facilitators be when delivering the manualised training? Interestingly, it was the facilitators (Limerick) and teachers from one study (Dublin 2) who appreciated the highly structured course; most other studies reported that some participating teachers felt flexibility was needed. Also, are there issues of acceptability and appropriateness that need to be addressed when transferring the program materials to new countries? Although it is standard practice to translate the language (e.g., English to Welsh), the creation of new video vignettes was regarded as worthwhile, because the standard American vignettes used in most studies were viewed as being at odds with non-American educational contexts. However, only one quantitative study (Jamaica) culturally adapted the video vignettes. This adaptation occurred after initial piloting work in which qualitative data indicated that teachers wanted to see Jamaican classrooms portrayed in IY TCM. It is unclear whether adapting the video vignettes resulted in better intervention effect sizes, as this was not tested. However, it is one example of improving acceptability of an intervention by taking into account stakeholders' feedback. With preliminary evidence for effectiveness in a number of countries,

the cultural adaptation of video vignettes would serve as a valuable next step in the process of potentially scaling up IY TCM implementation in these countries.

A number of child characteristics were identified in the qualitative strand as perceived moderators of IY TCM intervention effects. Notably, themes emerged around younger children and those in smaller classes responding better to IY TCM strategies than older children and those in larger classes. Also, children identified with SEN were reported as presenting additional difficulties for teachers and perhaps responding poorer to IY TCM strategies than their peers without identified SEN. With the exception of the North Carolina study, which conducted grade-level interactions for children's social, emotional, and academic outcomes, none of these possible child-level moderators was addressed in the quantitative evidence.

Two related themes that emerged in the qualitative strand related to the need for school leadership to be invested in IY TCM and for the whole school to be trained (rather than just one teacher). Neither of these constructs was addressed in quantitative analyses. No study measured the extent to which school leadership was invested in and supported IY TCM as a potential moderator (hypothesised to increase effectiveness). Similarly, no study implemented IY TCM across entire schools.

Discussion

Summary of main findings

Overall, results from the quantitative strand of this systematic review and multilevel meta-analysis suggest that IY TCM resulted in important changes at teacher and child levels. There were statistically significant moderate effects of IY TCM on reducing teacher negative classroom management strategies and high-risk child conduct problems. This suggests that IY TCM presents a valuable opportunity to reduce violence in schools. In particular, evidence from

the Jamaica study illustrates how, in settings where teachers' baseline levels of training are relatively low, large effects are possible in lessening the use of negative strategies (e.g., corporal punishment) and boosting the use of positive strategies (e.g., praise, establishing relationships). Similarly, in successfully reducing the amount of aggression and non-compliance among high-risk children in classrooms, all children will experience fewer violent outbursts from their peers. Additionally, a very large albeit non-significant effect of IY TCM on increasing teacher positive classroom management strategies was found, along with smaller, non-significant effects on children's prosocial behaviours. Where results were not statistically significant, patterns of effect sizes still trended towards the desired direction, and largely so in the case of teachers' positive strategies. Given how few included studies were involved in this multilevel meta-analysis, it is unsurprising that some confidence intervals crossed zero and some effects were not significant, even though these favoured intervention. The addition of more studies will serve to further refine these estimates. Child behaviour difficulties, a construct that incorporated a broader definition of problematic behaviour to include inattention/hyperactivity and peer problems, did produce statistically significant effects which followed the same patterns of effect sizes as all *a priori* outcomes, in that there was a small effect for all children, with the effect for high-risk children being three times greater. This is important because the qualitative data clearly showed that teachers' day-to-day concerns about behaviour include aggression/non-compliance *and* attention/peer relationships.

Findings from the qualitative strand of this systematic review and meta-synthesis of stakeholders' experiences suggest that IY TCM is acceptable and enjoyable for participants. The cyclical process of engaging in the learning process and materials individually as well as collectively produced a pattern of learning, hands-on practice, reflection, adjustments, and

further practice. Along the way, teachers felt as though they were building up a resource bank of strategies for future use. Stakeholders experienced numerous outcomes attributed to the training program, across teacher, classroom, and child levels. Despite the many positive views of IY TCM, there were critiques of the process and content. Most notably are the reservations to engage in role play, use of culturally discordant video vignettes, and perceived necessity of scaling-up training to the school level.

Quality of Evidence

None of the quantitative studies was judged to be at high risk of bias, and overall the quality of evidence included in this systematic review and multilevel meta-analysis of effectiveness trials was high. Vitally important for the quality of these results is the regular use of independent (blinded) observer-reported data on outcomes, rather than relying only on teacher-reported (non-blinded) data. The quality of evidence from the included qualitative studies was variable, and the largest barrier was a general lack of depth across all included studies. While the Limerick study consisted of multiple records including a PhD dissertation, resulting in a relatively rich description of the primary findings, and the Dublin 2 study was also reported in a PhD dissertation, other studies (e.g., Dublin 1, Wales) were extremely brief in their reporting and relied heavily on frequency counts and quasi-statistic descriptors (e.g., all, many, some) to present the data.

Potential Biases in the Review Process

The most notable limitation of this mixed methods review is the small number of included studies. Quantitatively, due to the small number of included studies ($n = 7$), we could not examine publication bias through a funnel plot; however, included studies were identified through the systematic database/website searches, and only one was found through contact with

experts in the field because it was in press at the time of searching (although its protocols were found via database searching). Due to having few IY TCM trials, the quantitative analyses are based on relatively few participants ($n_{teachers} = 336$; $n_{children} = 5,759$), with variation on exact numbers by outcome. Given the multilevel model used in the meta-analysis, having an increased sample size both at the individual participant level and the included study level would improve statistical power and minimize potential sampling bias.

Although the search terms were piloted and electronic databases were searched for ‘all fields’ rather than limited to keywords or titles/abstracts, it is possible that unpublished studies or reports not indexed in the main electronic databases could exist. However, there were no disagreements during the screening process between the two independent reviewers, and the very focused nature of this review on one branded intervention with clear criteria for study design made for a straightforward screening process.

Quantitatively, a comparison group meeting ‘treatment-as-usual’ is not a standardised control across included studies, unlike specifying IY TCM as the intervention. In conducting a multilevel meta-analysis that compares effect sizes calculated from standardised means, the analysis is in fact answering the question, ‘Is IY TCM more effective than treatment-as-usual?’ Without each included study explicating the components included in their respective control groups, it is not possible to establish how theoretically comparable the comparison conditions are. This limitation applies to all meta-analyses, but to the extent that authors do not over-generalise the implications that can be drawn from their studies this approach still provides us with the best information on program effectiveness.

Qualitatively, where contradictions existed within the synthesis (e.g., critiques section), these were highlighted and discussed. Often, however, studies did not contradict each other but

instead bias potentially crept in due to the predominant role the Limerick study played in the synthesis due to its richness (it served as the index paper because of its high quality and rich descriptions). The most robust sections of the synthesis are determined by convergence on the themes from multiple included studies. All of the teacher outcomes (knowledge, locus-of-control, emotional well-being, and teaching philosophy) were grounded in data that spanned the included studies. Similarly, the child outcomes of behaviour, social skills, and emotional well-being were robustly justified in the data across multiple included studies. The themes of role play/video vignettes and scaling-up implementation to the whole school emerged from data across the included studies, too. The classroom outcomes, remaining child outcomes, and other critiques are not as strongly substantiated in the available literature and should therefore be interpreted with caution. Despite these limitations, the qualitative meta-synthesis remains a valuable contribution because it is the first to examine the available experiential data on IY TCM and how new theory can emerge from understanding the collective findings from each primary study.

Agreements and Disagreements with Past Reviews

Overall, this current review and multilevel meta-analysis has produced results that further support the effectiveness of teacher-training as an intervention to benefit children's behaviour. Oliver and colleagues (2011) concluded that teacher classroom management programs do have a significant overall effect on decreasing problem behaviour in the classroom. Narrative synthesis of 14 included studies testing eight different types of interventions in the (Whear et al., 2013) review revealed inconsistent effects of teacher management programs on children's social and prosocial skills. There were trends towards decreased problem behaviours due to intervention,

and teacher positive management strategies were found to be moderately statistically significant in favour of intervention.

Implications for Practice and Policy

This review addressed one intervention that seeks to address the now global concern over children's mental health. In meta-analysing effects in a multilevel model and including data across five countries, there is clear evidence to support the use of IY TCM as an intervention to reduce violence in schools. This reduction is seen both from lessening teachers' use of negative strategies as well as in improvements in behaviour from the high-risk children in the classroom. Evidence comes primarily from high-income countries, but the Jamaica study demonstrates huge benefits can occur in low- and middle-income countries (LMICs). Practitioners and policy-makers can have increased confidence in these findings because these data incorporate both teacher-/parent-reports as well as independent observations, which have reduced bias because of blinding to treatment condition.

Results from the qualitative strand of this systematic review and qualitative meta-synthesis of stakeholders' experiences of IY TCM illustrate the complex learning process that teachers undergo and the benefits of this training as well as some critiques. Synthesizing this information adds valuable contextual information to the evidence-base for decision-makers, who can make more educated choices regarding whether the teachers in their school or area might benefit from embarking on such a journey and if so, how the training will be introduced, implemented, and measured. When considering taking a program like IY TCM to scale, extensive work in Wales provides key considerations for large-scale success, including having a local champion, linking the program into public policy, and rigorously evaluating the program in the new setting (Hutchings & Williams, 2017).

Concerns over (un)intentional adaptations also need to be addressed moving forward. Specifically, qualitative findings suggest that stakeholders might prefer adapted video vignettes when in non-American contexts. This would require an upfront expenditure of resources, although the idea is that were these adaptations to be carried out in form only (e.g. language translations, updated classrooms) and not in function (i.e., intended messages), then teachers might find the vignettes to be more acceptable and relevant to their own classroom practice. Those who might favour cultural adaptation of interventions recognise that adaptations can increase intervention acceptability and increase the effectiveness of the intervention among new groups (Lau, 2006). However, those who might be hesitant to adapt interventions cite apprehensions that stem from our lack of understanding about core components and the subsequent fear of omitting or changing core components (Elliott & Mihalic, 2004; Mihalic, 2002). Research by Leijten and colleagues (2016) showed that comparing transported parenting programs to homegrown parenting programs resulted in no statistically significant difference on effectiveness. This finding held across different brands, including the IY parenting program, and regions (including non-Western countries and LMICs).

Implications for Research

There is still a dearth of quantitative evidence on IY TCM, especially when compared to the corpus of evidence on the IY parenting program. Data is also relatively limited in its follow-up, with all studies measuring immediately post-intervention and some collecting follow-ups that are limited due to the wait-list design. This hinders our understanding of how sustainable post-intervention outcomes may or may not be. Such information is particularly important in the field of school-based interventions because a strong argument in favour of intervening via teachers is their universal reach to school-aged children and potential to influence the children in their

classrooms year after year. Future designs that allow for longer-term outcome measurements or phased implementation could start to address this gap. In particular, stepped-wedge designs where IY TCM is rolled out to schools over multiple years offer future researchers opportunities to experimentally test the effectiveness of whole-school implementation, dosage effects, as well as achieve longer-term outcomes.

A number of possible moderators and mediators were identified based on stakeholder experiences and explanations of how they viewed IY TCM as working (or not) for them. For example, among the included quantitative studies in this review, none addressed the possible moderating effect of children's SEN status on outcomes, even though this was identified in the qualitative data as a perceived concern by some teachers. Using qualitative evidence to drive explorations of potential moderators and mediators in future quantitative studies could provide deeper insights into for whom this training is effective, in which conditions, when, and through which mechanisms.

Conclusion

Teachers' concerns over problematic behaviours in the classroom have been long documented (Meister & Melnick, 2003; Veenman, 1984; Wickman, 1928), suggesting that this is a chronic problem that has not yet been properly addressed. Similarly, concerns over children's mental health are relatively stable across contexts and indicate that around 10 per cent of school-age children reach clinical cut-offs for diagnoses (H. Green, McGinnity, Meltzer, Ford, & Goodman, 2005; Merikangas et al., 2010; J. Murray et al., 2013; Wittchen et al., 2011) and another 20 per cent demonstrate problematic behaviours (J. Murray et al., 2013). By intervening at the teacher-level to change teachers' classroom management strategies, the aim is to address both the teacher's needs for increased skills in their profession and the children's social, emotional, and

behavioural needs. The IY TCM program is just one example of a classroom management intervention designed to do just that. Potentially, this provides a scalable public health solution to problems of wide concern to society, as well as to the education system. The results of this mixed methods systematic review of IY TCM suggest that it appears to be acceptable and well-received by teachers in addition to being effective for both child and teacher outcomes.

This review highlighted remaining gaps in the IY TCM evidence-base, particularly with regards to understanding the mechanisms of intervention and illuminating any moderators (e.g., SEN, age, class size) so that practitioners and policy-makers are best informed as to the effects they can expect to see if they were to select IY TCM for use in their area. By approaching this review from a mixed methods perspective, outcome and process data were able to be considered side-by-side. Future reviews should apply mixed methods when systematically reviewing interventions, as the results offer insights beyond any understanding that is achieved through examining effectiveness alone.

References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, D.C.
- Baker-Henningham, H. (2011). Transporting evidence-based interventions across cultures: Using focus groups with teachers and parents of pre-school children to inform the implementation of the Incredible Years Teacher Training Programme in Jamaica. *Child: Care, Health and Development*, 37(5), 649–661. <https://doi.org/10.1111/j.1365-2214.2011.01208.x>
- Baker-Henningham, H., Scott, S., Jones, K., & Walker, S. (2012). Reducing child conduct problems and promoting social skills in a middle-income country: Cluster randomised controlled trial. *The British Journal of Psychiatry*, 201(2), 101–108. <https://doi.org/10.1192/bjp.bp.111.096834>
- Baker-Henningham, H., & Walker, S. (2009). A qualitative study of teacher's perceptions of an intervention to prevent conduct problems in Jamaican pre-schools. *Child: Care, Health and Development*, 35(5), 632–642. <https://doi.org/10.1111/j.1365-2214.2009.00996.x>
- Baker-Henningham, H., & Walker, S. (2018). Effect of transporting an evidence-based, violence prevention intervention to Jamaican preschools on teacher and class-wide child behaviour: A cluster randomised trial. *Global Mental Health*, 5(e7), 1–16. <https://doi.org/10.1017/gmh.2017.29>
- Baker-Henningham, H., Walker, S., Powell, C., & Gardner, J. M. (2009a). A pilot study of the Incredible Years Teacher Training programme and a curriculum unit on social and emotional skills in community pre-schools in Jamaica. *Child: Care, Health and Development*, 35(5), 624–631. <https://doi.org/10.1111/j.1365-2214.2009.00964.x>
- Baker-Henningham, H., Walker, S., Powell, C., & Gardner, J. M. (2009b). Preventing behaviour

- problems through a universal intervention in Jamaican basic schools: A pilot study. *West Indian Medical Journal*, 58(5), 460–464. Retrieved from http://caribbean.scielo.org/scielo.php?script=sci_arttext&pid=S0043-31442009000500012
- Bandura, A. (1978). Social learning theory of aggression. *Journal of Communication*, 28(3), 12–29.
- Bonell, C., Jamal, F., Harden, A., Wells, H., Parry, W., Fletcher, A., ... Moore, L. (2013). Systematic review of the effects of schools and school environment interventions on health: evidence mapping and synthesis. *Public Health Research*, 1(1), 1–320. <https://doi.org/10.3310/phr01010>
- Bowlby, J. (1973). *Attachment and loss*. New York, NY: Basic Books.
- Carlson, J. S., Tired, H. B., Bender, S. L., & Benson, L. (2011). The influence of group training in the Incredible Years Teacher Classroom Management program on preschool teachers' classroom management strategies. *Journal of Applied School Psychology*, 27(2), 134–154. <https://doi.org/10.1080/15377903.2011.565277>
- Conduct Problems Prevention Research Group. (1995). Teacher - Social Competence Scale.
- Corbin, J., & Strauss, A. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3rd ed.). Thousand Oaks, CA: SAGE Publications, Inc.
- Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research* (2nd ed.). Thousand Oaks, CA: SAGE Publications, Inc.
- Davenport, J., & Tansey, A. (2009). *Outcomes of an Incredible Years Classroom Management Programme with Teachers from Multiple Schools*. Dublin, Ireland.
- Elliott, 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>

Evertson, C. M., & Weinstein, C. S. (2013). *Handbook of classroom management: Research, practice, and contemporary issues*. London, UK: Routledge.

Eyberg, S. M., & Robinson, E. A. (1981). Dyadic parent-child interaction coding system.

Fergusson, D. M., Horwood, L. J., & Stanley, L. (2013). A preliminary evaluation of the Incredible Years teacher programme. *New Zealand Journal of Psychology*, 42(2), 51–56.

Flemming, K. (2010). Synthesis of quantitative and qualitative research: An example using Critical Interpretive Synthesis. *Journal of Advanced Nursing*, 66(1), 201–217.

<https://doi.org/10.1111/j.1365-2648.2009.05173.x>

Ford, T., Hayes, R., Byford, S., Edwards, V., Fletcher, M., Logan, S., ... Ukoumunne, O. C. (2018). The effectiveness and cost-effectiveness of the Incredible Years® Teacher Classroom Management programme in primary school children: results of the STARS cluster randomised controlled trial. *Psychological Medicine*, 1–15.

<https://doi.org/10.1017/S0033291718001484>

Fossum, S., Handegård, B. H., & Britt Drugli, M. (2017). The Incredible Years Teacher Classroom Management programme in kindergartens: Effects of a universal preventive effort. *Journal of Child and Family Studies*, 26(8), 2215–2223.

<https://doi.org/10.1007/s10826-017-0727-3>

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative inquiry*. Hawthorne, NY: Aldine de Gruyter.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, 38(5), 581–586. <https://doi.org/10.1111/j.1469-7610.1997.tb01545.x>

- Green, H., McGinnity, Á., Meltzer, H., Ford, T., & Goodman, R. (2005). *Mental health of children and young people in Great Britain, 2004*. Basingstoke: Palgrave Macmillan.
- Green, J., & Thorogood, N. (2014). *Qualitative methods for health research* (3rd ed.). London, UK: SAGE Publications, Inc.
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. London, UK: Academic Press Inc.
- Herman, K. C., & Reinke, W. M. (2017). Improving teacher perceptions of parent involvement patterns: Findings from a group randomized trial. *School Psychology Quarterly*, 32(1), 89–104. <https://doi.org/10.1037/spq0000169>
- Hickey, G., McGilloway, S., Hyland, L., Leckey, Y., Kelly, P., Bywater, T. J., ... O'Neill, D. (2015). Exploring the effects of a universal classroom management training programme on teacher and child behaviour: A group randomised controlled trial and cost analysis. *Journal of Early Childhood Research*, 1–21. <https://doi.org/10.1177/1476718X15579747>
- Higgins, J. P. T., & Green, S. (2008). *Cochrane handbook for systematic reviews of interventions* (5th ed.). Wiley Online Library.
- Hutchings, J. (2012). *Support for teachers around the world*.
- Hutchings, J., Daley, D., Jones, K., Martin, P. A., Bywater, T. J., & Gwyn, R. (2007). Early results from developing and researching the Webster-Stratton Incredible Years Teacher Classroom Management training programme in North West Wales. *Journal of Children's Services*, 2(3), 15–26.
- Hutchings, J., Martin-Forbes, P., Daley, D., & Williams, M. E. (2013). A randomized controlled trial of the impact of a teacher classroom management program on the classroom behavior of children with and without behavior problems. *Journal of School Psychology*, 51(5), 571–

585. <https://doi.org/10.1016/j.jsp.2013.08.001>

Hutchings, J., & Williams, M. E. (2017). Taking the Incredible Years Child and Teacher Programs to scale in Wales. *Childhood Education, 93*(1), 20–28.

<https://doi.org/10.1080/00094056.2017.1275233>

Hyland, L. (2014). *The Incredible Years Teacher Classroom Management programme in Ireland: A process evaluation and observational assessment of teacher-pupil outcomes*. National University of Ireland - Maynooth.

Kam, C.-M., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based. *Prevention Science, 4*(1), 55–63.

Kennedy, Y. (2016). *Teacher perceptions of change through participation in the Incredible Years Teacher Classroom Management Programme*. University College London.

Korpershoek, H., Harms, T., De Boer, H., Van Kuijk, M., & Doolaard, S. (2016). A Meta-Analysis of the Effects of Classroom Management Strategies and Classroom Management Programs on Students Academic, Behavioral, Emotional, and Motivational Outcomes. *Review of Educational Research, 86*(3), 643–680.

<https://doi.org/10.3102/0034654315626799>

Lau, A. S. (2006). Making the case for selective and directed cultural adaptations of evidence-based treatments: Examples from parent training. *Clinical Psychology: Science and Practice, 13*(4), 295–310. <https://doi.org/10.1111/j.1468-2850.2006.00042.x>

Leckey, Y., Hyland, L., Hickey, G., Lodge, A., Kelly, P., Bywater, T., ... McGilloway, S. (2016). A mixed-methods evaluation of the longer-term implementation and utility of a teacher classroom management training programme in Irish primary schools. *Irish Educational Studies, 35*(1), 35–55. <https://doi.org/10.1080/03323315.2016.1147974>

- 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 and Adolescent Psychiatry, 55*(7), 610–617. <https://doi.org/10.1016/j.jaac.2016.05.003>
- Marlow, R., Hansford, L., Edwards, V., Ukoumunne, O., Norman, S., Ingarfield, S., ... Ford, T. (2015). Teaching classroom management - a potential public health intervention? *Health Education, 115*(3/4), 230–248. <https://doi.org/http://dx.doi.org/10.1108/HE-03-2014-0030>
- Martin, P. A. (2009). *From small acorns: The positive impact of adopting simple teacher classroom management strategies on global classroom behaviour and teacher-pupil relationships*. Bangor University. <https://doi.org/10.13140/RG.2.1.1168.7842>
- Martin, P. A., Daley, D., Hutchings, J., Jones, K., Eames, C., & Whitaker, C. J. (2010). Development and testing of a new classroom observation measure. *School Psychology International, 31*(3), 229–249. <https://doi.org/10.1177/0143034310362040>
- McGilloway, S., Hyland, L., Ní Mháille, G., Lodge, A., O'Neill, D., Kelly, P., ... Donnelly, M. (2010). *Positive classrooms, positive children: A randomised controlled trial to investigate the effectiveness of the Incredible Years Teacher Classroom Management programme in an Irish context*. Ireland. Retrieved from www.iyirelandstudy.ie
- McGilloway, S., Ní Mháille, G., Furlong, M., Hyland, L., Leckey, Y., Kelly, P., ... Donnelly, M. (2012). *Parents, teachers, and early childhood intervention: Long-term outcomes of the Incredible Years Parent and Teacher Classroom Management training programmes (Combined 12-month report)*. Dublin, Ireland. Retrieved from www.iyirelandstudy.ie
- Meister, D. G., & Melnick, S. A. (2003). National new teacher study: Beginning teachers' concerns. *Action in Teacher Education, 24*(4), 87–94.

<https://doi.org/10.1080/01626620.2003.10463283>

Merikangas, K. R., He, J. P., Brody, D., Fisher, P. W., Bourdon, K., & Koretz, D. S. (2010).

Prevalence and Treatment of Mental Disorders Among US Children in the 2001-2004

NHANES. *Pediatrics*, *125*(1), 75–81. <https://doi.org/10.1542/peds.2008-2598>

Mihalic, S. (2002). The Important of Implementation Fidelity. *Journal of Chemical Information and Modeling*, 1–16. <https://doi.org/10.1017/CBO9781107415324.004>

Murray, D., Murr, N., & Rabiner, D. L. (2012). Preliminary effects of the Incredible Years

Teacher training program on classroom management skills. In *Paper presented at Society for Research on Educational Effectiveness* (pp. 1–11).

Murray, D., Rabiner, D. L., & Carrig, M. M. (2014). Grade level effects of the Incredible Years

Teacher training program. In *Paper presented at Society for Research on Educational Effectiveness* (pp. 1–10).

Murray, D. W., Rabiner, D. L., Kuhn, L., Pan, Y., & Sabet, R. F. (2018). Investigating teacher and student effects of the Incredible Years Classroom Management Program in early elementary school. *Journal of School Psychology*, *67*, 119–133.

<https://doi.org/10.1016/j.jsp.2017.10.004>

Murray, J., Anselmi, L., Gallo, E. A. G., Fleitlich-Bilyk, B., & Bordin, I. A. (2013).

Epidemiology of childhood conduct problems in Brazil: Systematic review and meta-analysis. *Social Psychiatry and Psychiatric Epidemiology*, *48*(10), 1527–1538.

<https://doi.org/10.1007/s00127-013-0695-x>

Nye, E., Melendez-Torres, G. J., & Bonell, C. (2016). Origins, methods and advances in qualitative meta-synthesis. *Review of Education*, *4*(1), 57–79.

<https://doi.org/10.1002/rev3.3065>

- Oliver, R. M., Wehby, J. H., & Reschly, D. J. (2011). Teacher classroom management practices: effects on disruptive or aggressive student behavior. *Campbell Systematic Reviews*, 4, 1–55.
- Patterson, G. R., DeBaryshe, B., & Ramsey, E. (1990). A developmental perspective on antisocial behavior. *American Psychologist*, 44, 329–335.
- Pidano, A. E., & Allen, A. R. (2015). The Incredible Years Series: A review of the independent research base. *Journal of Child and Family Studies*, 24(7), 1898–1916.
<https://doi.org/10.1007/s10826-014-9991-7>
- Polanin, J. R. (2013). *Addressing the issue of meta-analysis multiplicity in education and psychology*. Loyola University Chicago.
- Reinke, W. M., Herman, K. C., & Dong, N. (2014). The Incredible Year Teacher Classroom Management program: Initial findings from a group randomized control trial. In *Paper presented at Society for Research on Educational Effectiveness* (pp. 1–7).
- Reinke, W. M., Herman, K. C., & Dong, N. (2016). *The Incredible Years Teacher Classroom Management program: Outcomes from a group randomized trial*. <https://doi.org/Manuscript> submitted for publication. Retrieved from <http://www.incredibleyears.com/wp-content/uploads/Reinke-IY-TCM-Program-Outcomes.pdf>
- Shepard, S. A., Doctoroff, G., Dickstein, S., & Seifer, R. (2008). An independent replication of the Incredible Years Classroom Management Teacher Training program in Head Start. In *Poster presented at the Society for Prevention Research*.
- Snyder, H. (2001). Epidemiology of official offending. In R. Loeber & D. P. Farrington (Eds.), *Child delinquents: Development, intervention and service needs* (pp. 25–46). Thousand Oaks, CA: SAGE Publications, Inc.
- Tapp, J., Wehby, J. H., & Ellis, D. (1995). A multiple option observation system for

- experimental studies: MOOSES. *Behavior Research Methods, Instruments, & Computers*, 27(1), 25–31. <https://doi.org/10.3758/BF03203616>
- Teddlie, C., & Tashakkori, A. (2009). *The foundations of mixed methods research: Integrating quantitative and qualitative techniques in the social and behavioral sciences*. Thousand Oaks, CA: SAGE Publications, Inc.
- United Nations General Assembly. (2015). Resolution adopted by the General Assembly on 25 September 2015. Washington, D.C.: United Nations.
- Van Den Noortgate, W., López-López, J. A., Marín-Martínez, F., Sánchez-Meca, J., & Van Den Noortgate, W. (2013). Three-level meta-analysis of dependent effect sizes. *Behav Res*, 45, 576–594. <https://doi.org/10.3758/s13428-012-0261-6>
- Van Den Noortgate, W., & Onghena, P. (2003). Multilevel meta-analysis: A comparison with traditional meta-analytical procedures. *Educational and Psychological Measurement*, 63(5), 765–790. <https://doi.org/10.1177/0013164402250127>
- Veenman, S. (1984). Perceived problems of beginning teachers. *Review of Educational Research*, 54(2), 143–178.
- Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software*, 36(3), 1–48. <https://doi.org/10.18637/jss.v036.i03>
- 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>
- Webster-Stratton, C. (2005). Teacher Classroom Management Strategies Questionnaire. Retrieved from <http://www.incredibleyears.com/for-researchers/measures/>
- Webster-Stratton, C., Gaspar, M. F., & Seabra-Santos, M. J. (2012). Incredible Years® Parent,

Teachers and Children's Series: Transportability to Portugal of Early Intervention Programs for Preventing Conduct Problems and Promoting Social and Emotional Competence.

Psychosocial Intervention, 21(2), 157–169. <https://doi.org/10.5093/in2012a15>

Webster-Stratton, C., Reid, M. J., & Hammond, M. (2001). Preventing conduct problems, promoting social competence: A parent and teacher training partnership in Head Start.

Journal of Clinical Child & Adolescent Psychology, 30(3), 283–302.

Whear, R., Thompson-Coon, J., Boddy, K., Ford, T., Racey, D., & Stein, K. (2013). The effect of teacher-led interventions on social and emotional behaviour in primary school children: a systematic review. *British Educational Research Journal*, 39(2), 383–420.

<https://doi.org/10.1080/01411926.2011.650680>

Wickman, E. K. (1928). *Children's behavior and teachers' attitudes*. New York, NY: The Commonwealth Fund.

Wittchen, H. U., Jacobi, F., Rehm, J., Gustavsson, A., Svensson, M., Jönsson, B., ...

Steinhausen, H. C. (2011). The size and burden of mental disorders and other disorders of the brain in Europe 2010. *European Neuropsychopharmacology*, 21(9), 655–679.

<https://doi.org/10.1016/j.euroneuro.2011.07.018>

World Health Organization. (1992). *The ICD-10 classification of mental and behavioural disorders: Clinical descriptions and diagnostic guidelines*. Geneva: World Health Organization.

| Table 1 | | |
|------------------------|--|------|
| <i>Search strategy</i> | | |
| Databases searched: | 1. Applied Social Sciences Index and Abstracts (ASSIA) | 1320 |
| | 2. British Education Index | 12 |
| | 3. Cochrane Central Register of Controlled Trials (CENTRAL) | 98 |
| | 4. Cumulative Index to Nursing and Allied Health Literature (CINAHL) | 82 |
| | 5. EMBASE | 149 |
| | 6. Education Resources Information Center (ERIC) | 95 |
| | 7. Linguistics and Language Behaviour Abstracts | 234 |
| | 8. MEDLINE | 131 |
| | 9. PAIS International | 1033 |
| | 10. ProQuest Dissertations and Theses | 178 |
| | 11. PsycINFO | 1612 |
| | 12. Social Services Abstracts | 400 |
| | 13. Sociological Abstracts | 1697 |
| Websites searched: | 1. Campbell Collaboration Library https://www.campbellcollaboration.org/library.html | 0 |
| | 2. Cochrane Collaboration Review Library http://www.cochranelibrary.com/ | 8 |
| | 3. Incredible Years Series Library http://www.incredibleyears.com/research-library/ | 102 |
| | 4. Centre for Evidence-Based Early Intervention https://www.bangor.ac.uk/psychology/cebei/publications.php.en | 105 |
| Search terms: | 1. “incredible years”.af. | |
| | 2. “program*”.af. | |
| | 3. “teacher classroom management”.af. | |
| | 4. “teacher training”.af. | |
| | 5. 2 or 3 or 4 | |
| | 6. 1 and 5 | |

Table 2

Characteristics of included studies in the quantitative strand

| Study ID | Participants | Intervention and control | Outcomes | Analysis | Risk of bias assessment |
|---|---|--|---|--|--|
| England (Ford et al., 2018) | 80 teachers from 80 schools 2075 children | IY TCM (6 sessions) Wait-list, treatment-as-usual control | Teacher outcomes Teacher-reported classroom management strategies, satisfaction with IY TCM, burnout, well-being, and self-efficacy Child outcomes Teacher- and parent-reported child behaviour difficulties and prosocial behaviour, child-reported feelings about school, teacher-reported child academics | Child outcomes Random effects linear regression models using complete case analysis; analyses pre-specified and independently reviewed | Recruitment bias: Low risk Selection bias: Low risk Performance bias: Unclear risk Detection bias: Unclear risk Attrition bias: Low risk Statistical analysis bias: Low risk Reporting bias: Low risk |
| Jamaica (Baker-Henningham et al., 2012; Baker-Henningham & Walker, 2018) | 24 community pre-schools (urban) 3 children per class (high-risk only) | IY TCM (8 sessions; Culturally adapted to Jamaican context from pilot study in Baker-Henningham et al., 2012) Treatment-as-usual (Regular, Ministry of Education in-service workshops; Received | Child outcomes Observed child behaviour Teacher-reported child behaviour Parent-reported child behaviour Child attendance Parent outcomes Self-reported attitude to school | Child outcomes Multilevel regression models for continuous data; Random-effects logistical regression models estimated with Markov chain Monte-Carlo methods for binary data | Recruitment bias: Low risk Selection bias: Low risk Performance bias: Unclear risk Detection bias: Unclear risk Attrition bias: Low risk Statistical analysis bias: Low risk |

| | | identical educational materials) | | | Reporting bias: Low/Unclear risk |
|--|--|--|---|---|--|
| Limerick, Ireland (Hickey et al., 2015; Leckey et al., 2016; McGilloway et al., 2010) | 11 primary schools (8 urban, 2 semi-urban, 1 rural) 22 teachers (2 per school) 12 children per class | IY TCM (5 sessions) Wait-list control | Teacher outcomes Self-reported use of positive and negative strategies (frequency) and perceived usefulness of these strategies Observed positive strategies, praise, negative strategies, indirect commands, direct commands, and no opportunity for child to respond Child outcomes Teacher-reported child behaviour difficulties and prosocial behaviours Observed negative behaviour, positive behaviour, compliance, and non-compliance | Teacher outcomes ANCOVA Child outcomes Robust multiple linear regression | Recruitment bias: Low risk Selection bias: Low risk Performance bias: Unclear risk Detection bias: Unclear risk Attrition bias: Low risk Statistical analysis bias: Low/Unclear risk Reporting bias: Low risk |
| Missouri, US (Herman & Reinke, 2017; Reinke et | 105 teachers from 9 urban schools 1818 children | IY TCM (6 sessions) Wait-list, treatment-as-usual control | Child outcomes Administrative data on child demographics, teacher-reported child academic competence, standardised tests | Child outcomes Three-level hierarchical linear modelling, with multiple imputation | Recruitment bias: Low risk Selection bias: Low risk |

| | | | | | |
|---|---|--|---|---|---|
| al., 2014, 2016) | | | of reading and mathematics, observations and teacher-reports of child behaviour | using Markov chain Monte Carlo method for missing data | <p>Performance bias: Unclear risk</p> <p>Detection bias: Unclear risk</p> <p>Attrition bias: Low risk</p> <p>Statistical analysis bias: Low risk</p> <p>Reporting bias: Low risk</p> |
| | | | <p>Parent outcomes Teacher-reported parent involvement</p> | | |
| North Carolina, US (D. Murray et al., 2012, 2014; D. W. Murray et al., 2018) | 91 teachers from 11 rural and semi-rural schools 1192 children | IY TCM (5 sessions) Wait-list, treatment-as-usual control | <p>Teacher outcomes Observed teacher management strategies, teacher-reported satisfaction</p> <p>Classroom outcomes Observed classroom environment</p> <p>Child outcomes Teacher-reported child social-emotional outcomes, child behavioural and attention outcomes, and child academic outcomes in literacy/reading and mathematics</p> | <p>Child outcomes Two- and three-level hierarchical linear modelling, tested multiple imputations but did not impact results so analyses using listwise deletion</p> | <p>Recruitment bias: Low risk</p> <p>Selection bias: Low risk</p> <p>Performance bias: Unclear risk</p> <p>Detection bias: Unclear risk</p> <p>Attrition bias: Low risk</p> <p>Statistical analysis bias: Low risk</p> <p>Reporting bias: Low risk</p> |
| Rhode Island, US | 13 lead teachers | IY TCM (6 sessions) | <p>Teacher outcomes Teacher-reported positive and</p> | Not addressed on poster | <p>Recruitment bias: Unclear risk</p> |

| | | | | | |
|--|---|---|---|-------------------------------|--|
| (Shepard et al., 2008) | 13 teacher assistants 231 children | Treatment-as-usual (Standard Head Start training) | negative strategies, approach with parents, and relationship with children Classroom outcomes Observed classroom environment Child outcomes Teacher-reported child social skills and child behaviour problems | | Selection bias: Unclear risk Performance bias: Unclear risk Detection bias: Unclear risk Attrition bias: Low/Unclear risk Statistical analysis bias: Unclear risk Reporting bias: Low/Unclear risk |
| Wales (Hutchings et al., 2013; Martin, 2009) | 12 teachers (8 full-time, 4 part-time) 107 children Welsh-speaking classrooms | IY TCM (5 sessions) Wait-list control (No in-service or other professional development training sessions offered during trial) | Teacher outcomes Observed teacher management strategies Teacher-reported satisfaction with program Child outcomes Observed child behaviours | Hierarchical linear modelling | Recruitment bias: Low risk Selection bias: Low risk Performance bias: Unclear risk Detection bias: Low risk Attrition bias: Low risk Statistical analysis bias: Low risk Reporting bias: Low risk |

| Table 3 | | | | | | | |
|--|-------|--------------|------|----------------|----------------------------|----------------------------|-------------------------------|
| <i>Results of multilevel meta-analysis</i> | | | | | | | |
| Primary outcomes | N(k) | β_{00} | SE | 95% CI | I ² | Tau ² | Chi ² (p-value) |
| Teacher negative strategies | 8(3) | -0.49* | 0.20 | (-0.89, -0.09) | L1: 0.49 L2: 0.00 | L1: 0.16 L2: 0.00 | 13.66 (0.06) |
| Teacher positive strategies | 11(4) | -0.73 | 0.61 | (-1.92, 0.45) | L1: 0.26 L2: 0.65 | L1: 0.44 L2: 1.10 | 79.66 (<0.0001) |
| Child conduct problems (All) | 12(4) | -0.05 | 0.04 | (-0.13, 0.03) | L1: 0.04 L2: 0.00 | L1: 0.00 L2: 0.00 | 10.27 (0.51) |
| Child conduct problems (High-risk) | 9(3) | -0.35*** | 0.07 | (-0.49, -0.21) | L1: 0.00 L2: 0.00 | L1: 0.00 L2: 0.00 | 5.29 (0.73) |
| Secondary outcomes | | | | | | | |
| Child prosocial behaviours (All) | 8(5) | -0.12 | 0.09 | (-0.30, 0.06) | L1: 0.11 L2: 0.26 | L1: 0.01 L2: 0.02 | 15.27 (0.03) |
| Child prosocial behaviours (High-risk) | 4(3) | -0.31 | 0.17 | (-0.65, 0.03) | L1: 0.40 L2: 0.00 | L1: 0.06 L2: 0.00 | 8.23 (0.04) |

*N=number of effect sizes (Level 1); k=number of studies (Level 2); β_{00} = Hedges' g ('Grand Mean'); * $p<0.05$, ** $p<0.01$, *** $p<0.001$*

Table 4

Characteristics of included studies in the qualitative strand

| Study ID | Participants | Method | Topics | Analysis | Quality assessment |
|---|--------------|---|--|--|---|
| Dublin, Ireland (1) (Davenport & Tansey, 2009) | 15 teachers | Written survey, post-intervention | Teacher surveys Confidence managing child behaviours, recommendation of IY TCM workshops to colleagues, endorsement of IY TCM content, perceived improvement in child behaviours, perceptions of course materials, timing of course, perceived usefulness of course strategies | Frequency counts based on topic guide headings | Rigor of sampling: Medium quality Rigor of data collected: Low quality Rigor of analysis: Low quality Findings grounded in data: Low quality Breadth and depth: Low quality Overall reliability and trustworthiness: Low quality Overall usefulness of findings for this review: Low quality |
| Dublin, Ireland (2) (Kennedy, 2016) | 12 teachers | 5 focus groups (1 each month during intervention) and individual semi-structured interviews 6-8 weeks post-intervention | Teacher perceptions of change due to IY TCM (at teacher and child levels), facilitators and barriers of change | Theoretical (deductive) thematic analysis | Rigor of sampling: High quality Rigor of data collected: High quality Rigor of analysis: Medium quality Findings grounded in data: High quality Breadth and depth: Medium quality |

| | | | | | Overall reliability and trustworthiness : Medium quality |
|--|--|---|---|---|--|
| | | | | | Overall usefulness of findings for this review: High quality |
| England (Marlow et al., 2015) | 31 teachers and 16 head-teachers from state-funded primary schools | Teachers Focus groups comprising intervention training group members, one-month post-intervention | Teacher focus groups Course content, delivery, theory, group learning, skills and materials, feasibility, dissemination, and research process | Thematic analysis and Framework approach (coding frameworks based on topic guides but also developed iteratively during analysis) | Rigor of sampling: High quality Rigor of data collected: Medium quality Rigor of analysis: Medium quality Findings grounded in data: High quality Breadth and depth: Medium quality Overall reliability and trustworthiness : Medium quality Overall usefulness of findings for this review: High quality |
| Jamaica (Baker-Henningham & Walker, 2009) | 15 teachers from 3 community pre-schools | One-to-one, in-person in-depth interviews | Post-intervention interviews Opinion of training program and perceptions of benefits, how teachers used the strategies | Framework approach | Rigor of sampling: High quality Rigor of data collected: Medium quality Rigor of analysis: Medium quality |

| | | | | | |
|---|--|--|--|--|--|
| | | | and most/least useful strategies, views on how intervention brought about changes, suggestions for future implementations | | <p>Findings grounded in data: High quality</p> <p>Breadth and depth: Medium quality</p> <p>Overall reliability and trustworthiness: High quality</p> <p>Overall usefulness of findings for this review: High quality</p> |
| Limerick, Ireland (Hyland, 2014; McGilloway et al., 2010, 2012) | <p>11 teachers (post-intervention) from 6 schools</p> <p>6 teachers (1-year post-intervention follow-up)</p> <p>Group facilitators</p> | One-to-one, in-person semi-structured interviews | <p>Post-intervention interviews</p> <p>Current classroom management issues, strategies used to deal with classroom difficulties, participation experiences, post-training classroom management</p> <p>1-year follow-up</p> <p>How do teachers' experiences change over time, perceived long term effectiveness, best aspects of program and any barriers</p> | Thematic analysis and Framework approach | <p>Rigor of sampling: High quality</p> <p>Rigor of data collected: High quality</p> <p>Rigor of analysis: High quality</p> <p>Findings grounded in data: High quality</p> <p>Breadth and depth: High quality</p> <p>Overall reliability and trustworthiness: High quality</p> <p>Overall usefulness of findings for this review: High quality</p> |

| | | | | | |
|--|---|--------------------------|---|---------------------------------|--|
| Wales (Hutchings et al., 2007) | 21 teachers (IY TCM courses had been delivered in both Welsh and English) | One-to-one interviews | Post- intervention interviews Background information, IY TCM course, classroom application of skills learned, practical implications, comments and suggestions for future | Thematic content analysis | Rigor of sampling: High quality Rigor of data collected: Unclear quality Rigor of analysis: Unclear quality Findings grounded in data: Medium quality Breadth and depth: Medium quality Overall reliability and trustworthiness : Medium quality Overall usefulness of findings for this review: Medium quality |
|--|---|--------------------------|---|---------------------------------|--|

| Table 5 | |
|--|--|
| <i>Illustrative quotations to support findings of the qualitative meta-synthesis</i> | |
| Qualitative meta-synthesis themes | Quotations |
| Cycle of learning | <p>“The way in which the groups were facilitated, the group size and the time for teachers to get to know each other over a longer period of time helped to create a mutually supportive, trusting environment. This was valued because teachers liked being able to share difficulties in a non-judgemental environment (reducing isolation) and they felt they could try out new techniques, fail and try again.” (Marlow et al., 2015, p. 242)</p> |
| Perceived IY TCM outcomes | <p>“They reported increased knowledge of young children’s needs and abilities in terms of a better understanding of child development, using more developmentally appropriate strategies, using more interactive teaching methods, spending individual time with children and showing children more affection. Thus, teachers believed they now had more appropriate expectations for child behaviour and better understanding of the large individual differences in children’s development and of the need to give attention and encouragement to all children.” (Baker-Henningham & Walker, 2009, p. 634)</p> <p>“Teachers also reported that children had shown improvements in their overall conduct within the classroom. Several different aspects of child behaviour were discussed by teachers as having noticeably changed from pre- to post-TCM. Children had developed better peer relations, they had become more pro-social, and showed a greater willingness to help and encourage each other. The modelling of positive behaviour by teachers was regarded as being instrumental in promoting pro-social behaviour in the class, even amongst children who had previously displayed good class behaviour.” (McGilloway et al., 2010, p. 19)</p> <p>“They consider that the programme has ‘trained’ them in a step-by-step manner to implement positive behaviour management strategies and has become a ‘way of life’.” (Kennedy, 2016, p. 87)</p> |
| Critiques of the program | <p>“Generally, teachers regard the IYTCM strategies to be of practical use in improving classroom discipline, although it was noted that for a minority of children, some techniques were ineffective. For example, the inclusion of children with special needs in the mainstream classroom and class size/age of pupils were seen as a challenge to implementation. Overall, teachers recognised that the programme worked, but not for every child.” (McGilloway et al., 2012, p. 30)</p> <p>“[...]the inconsistency of other staff was a barrier [...] There was a strong feeling among the teachers that other, non-teaching, staff in the school also needed training in behaviour management, with seven (33%) mentioning teaching assistants and others mentioning head teachers, playground and canteen staff.” (Hutchings et al., 2007, pp. 20-21)</p> |

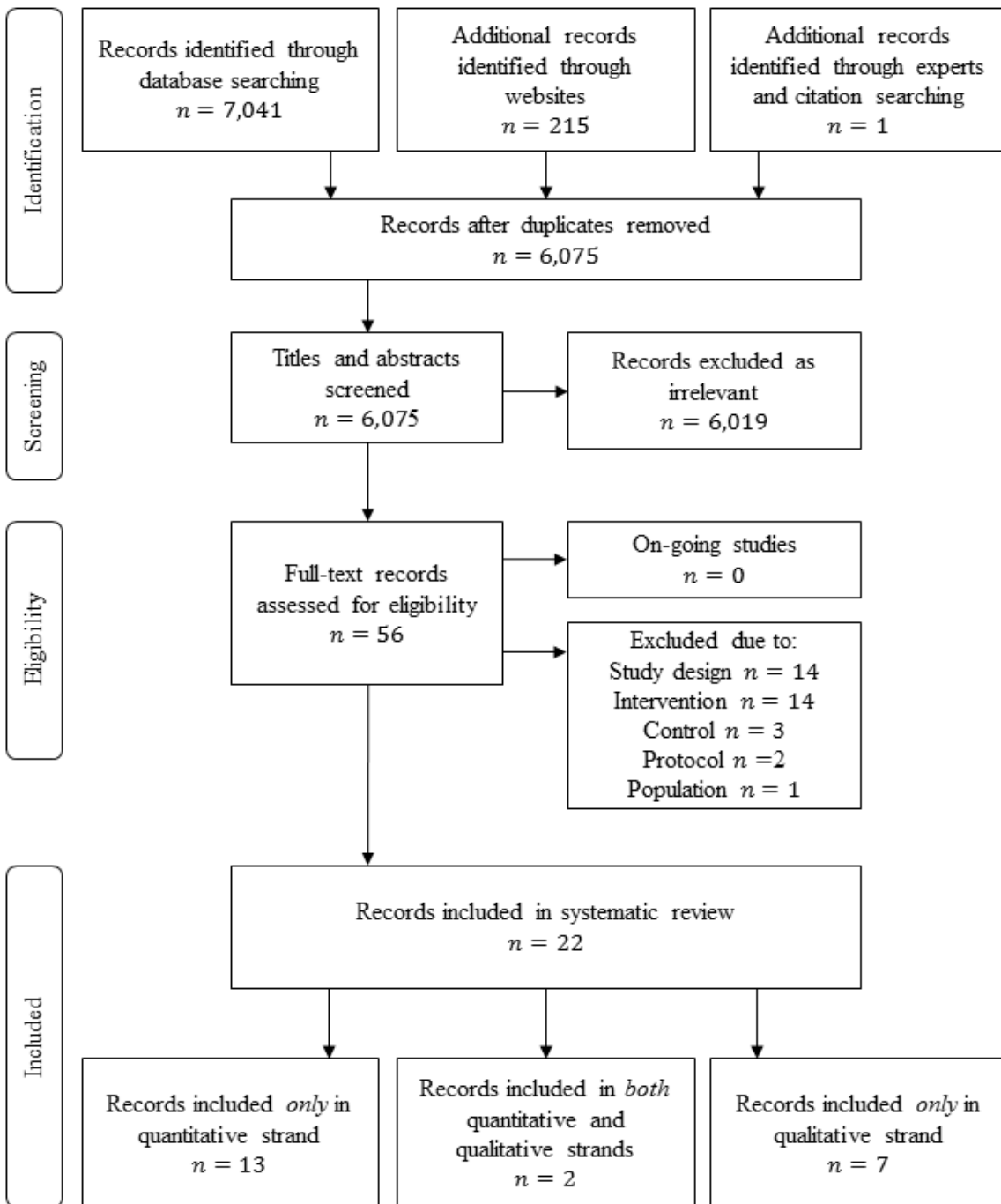


Figure 1. PRISMA flow chart of search results and study selection

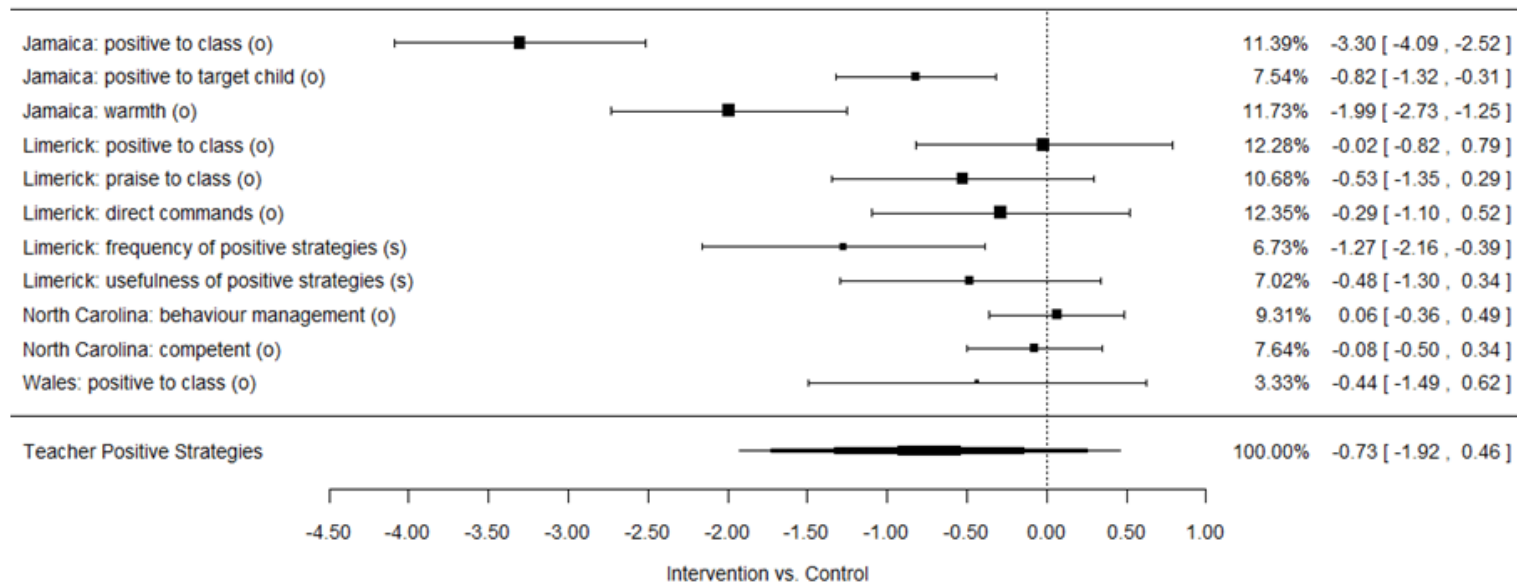
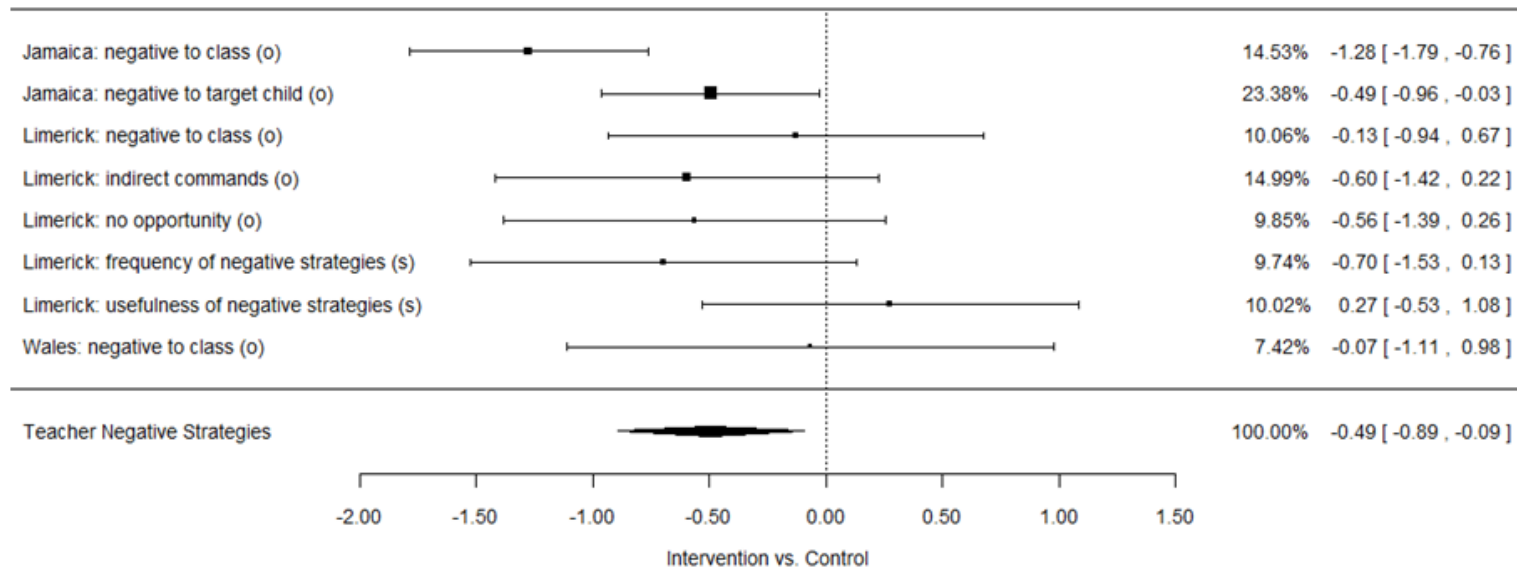


Figure 2. Multilevel meta-analysis of teacher strategies (o = observed; s = self-reported)

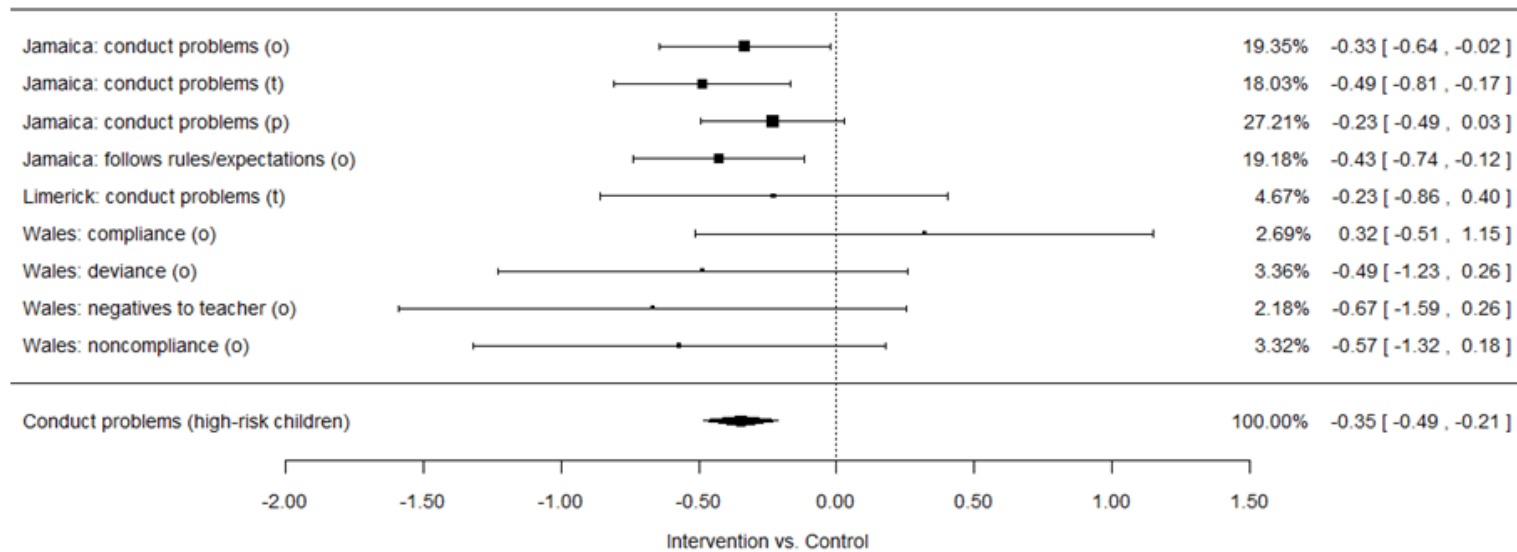
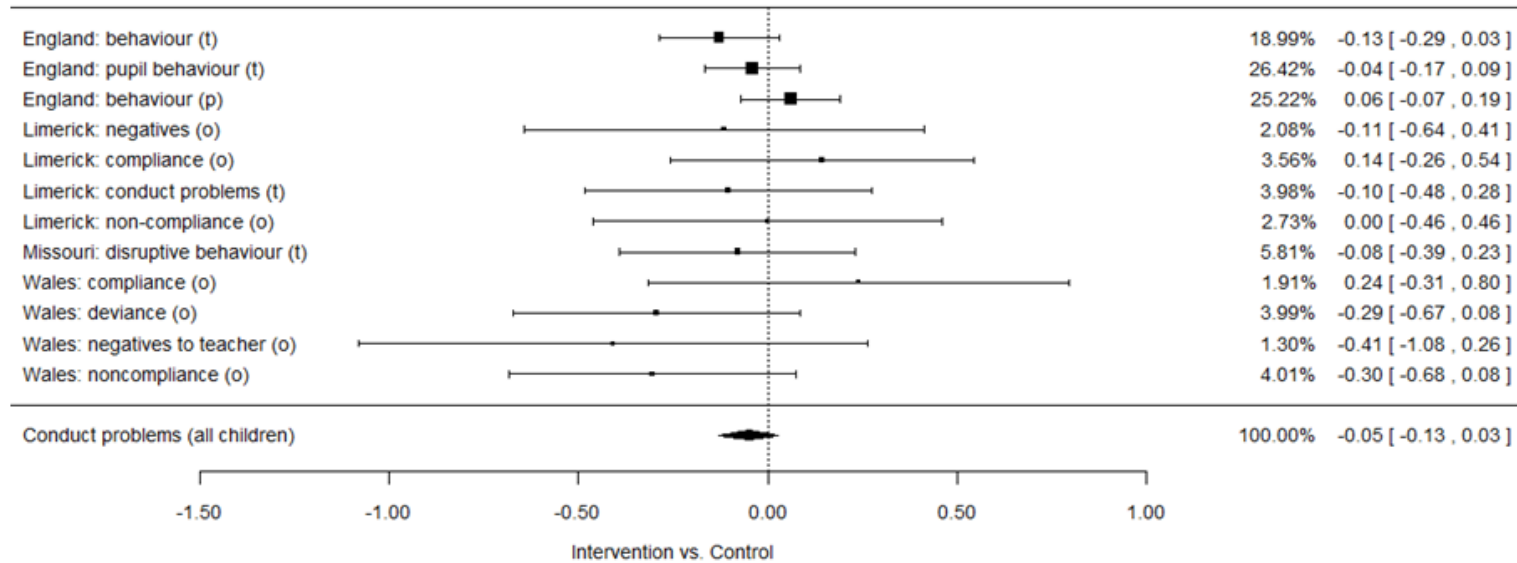


Figure 3. Multilevel meta-analysis of child conduct problems (t = teacher-reported, p = parent-reported, o = observed)

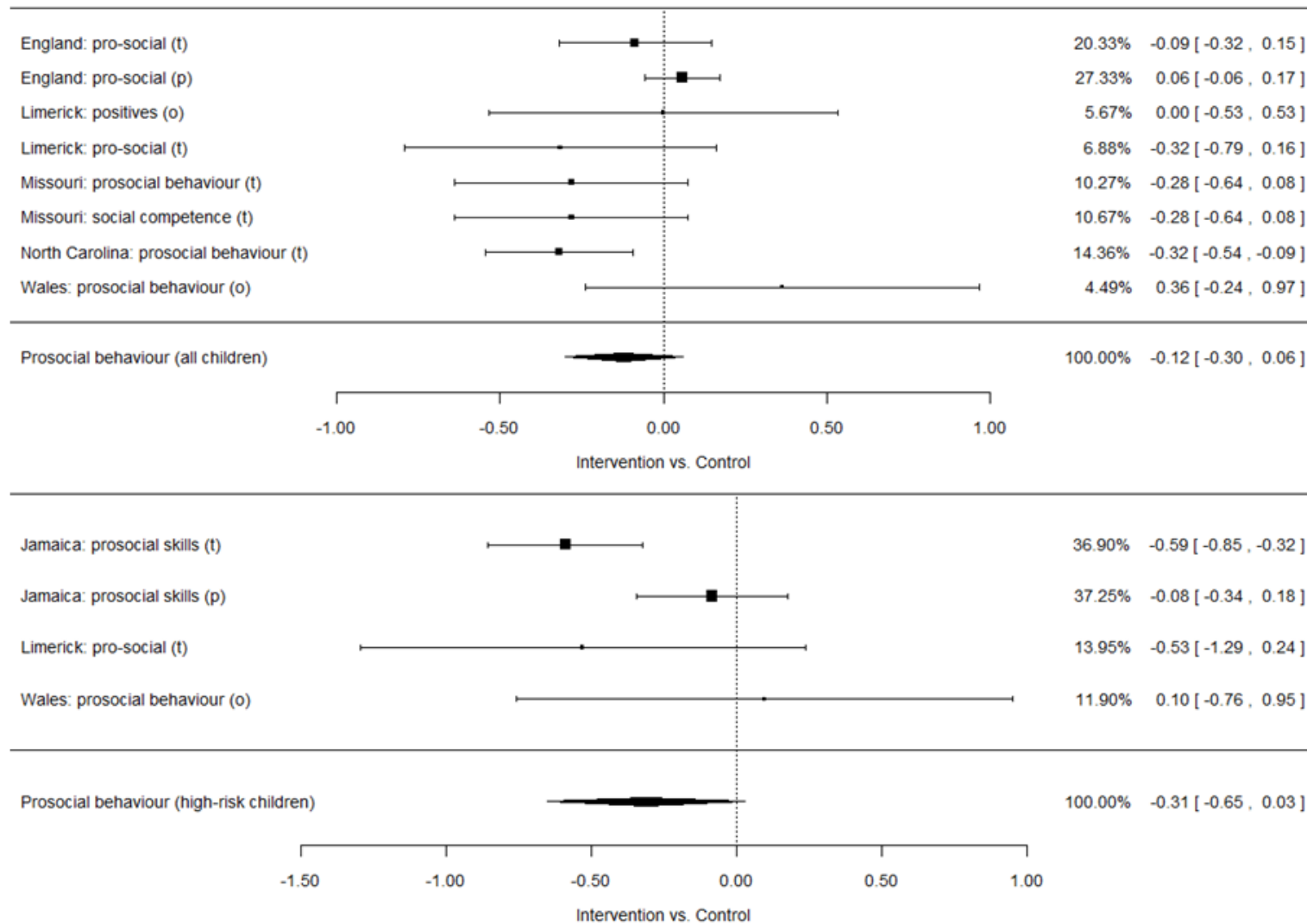


Figure 4. Multilevel meta-analysis of child prosocial behaviour (t = teacher-reported, p = parent-reported, o = observed)

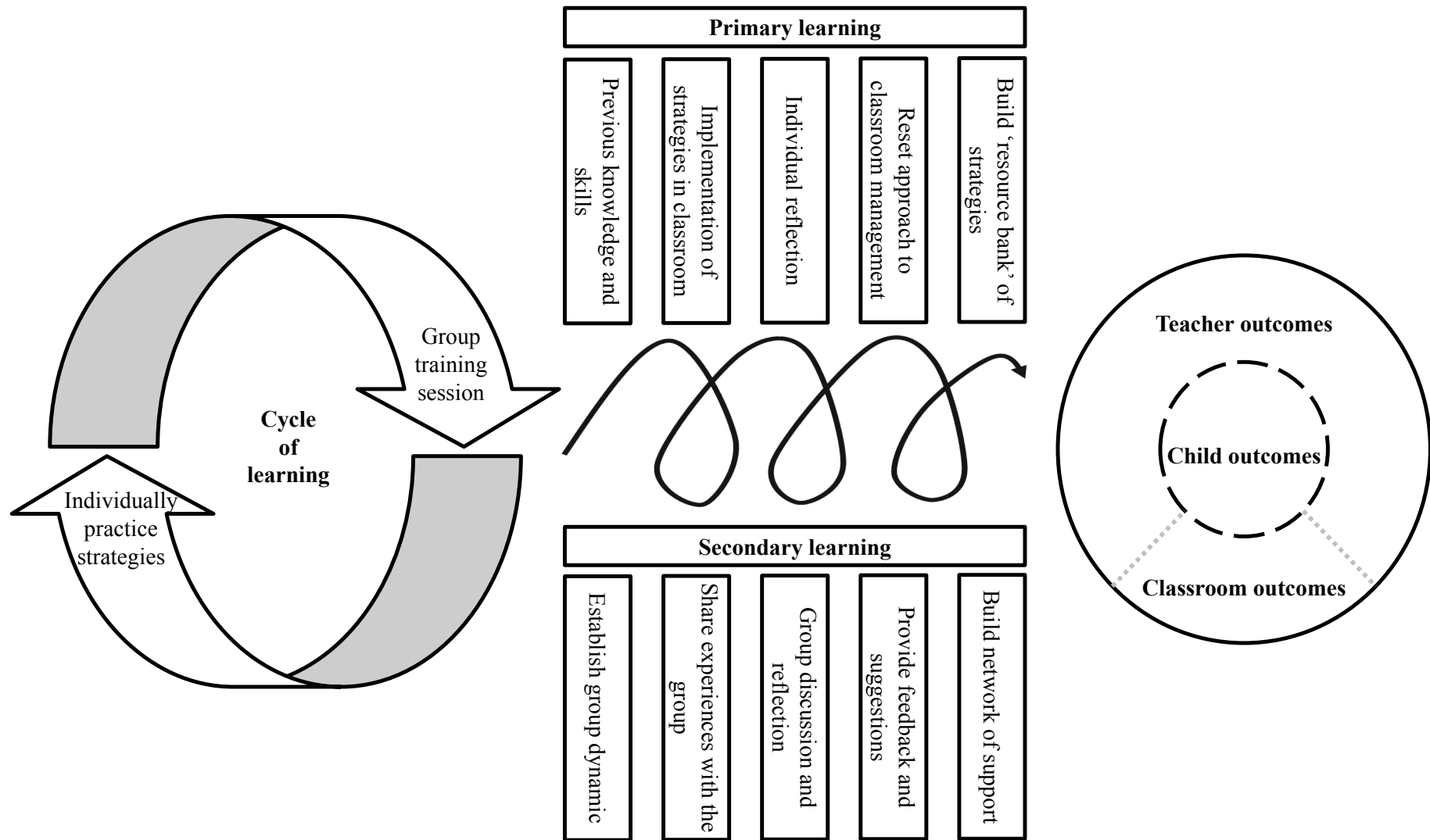


Figure 5. Qualitative meta-synthesis of experienced process participating in IY TCM course

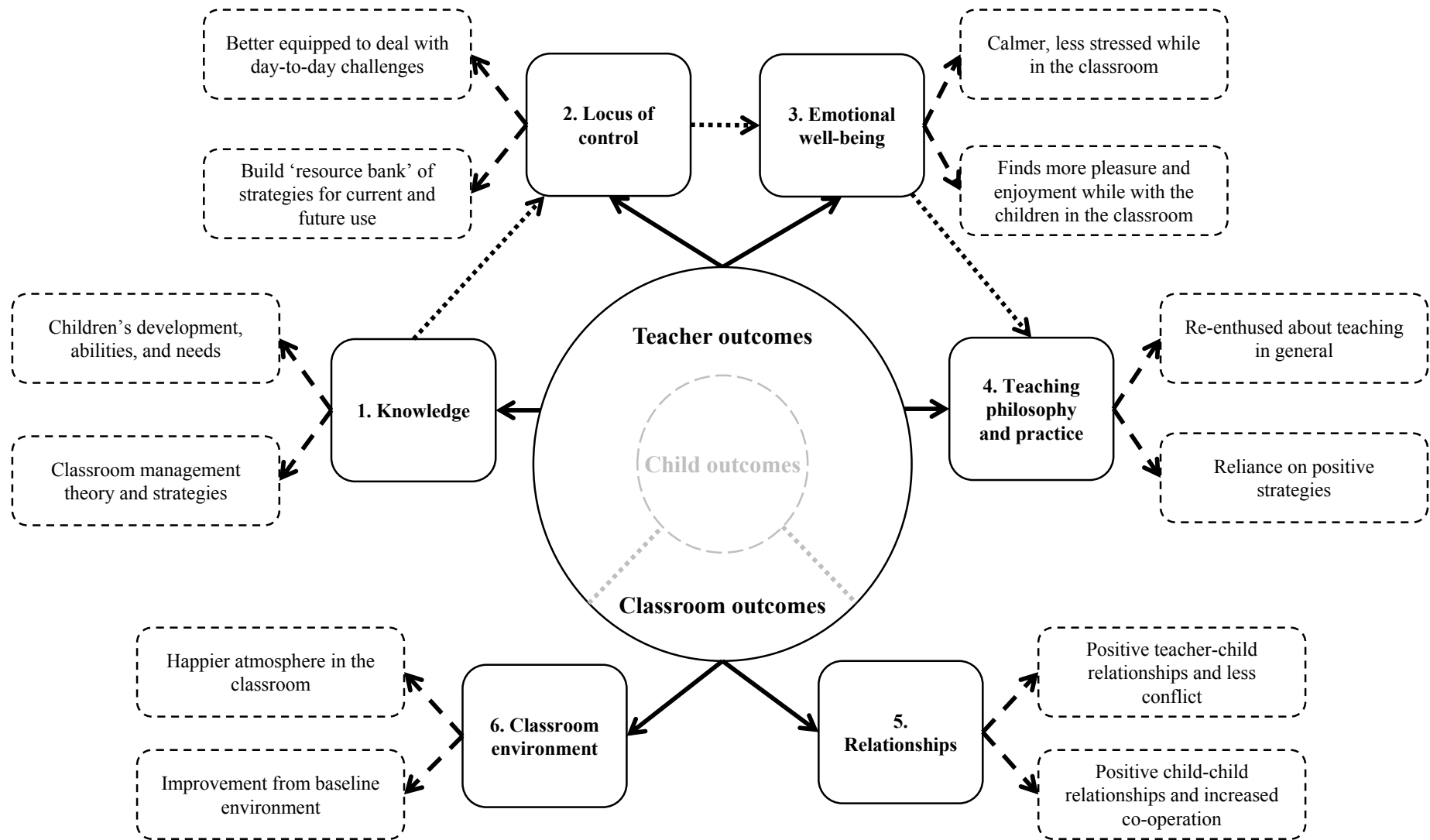


Figure 6. Qualitative meta-synthesis of experienced teacher outcomes