

Case Study 1: An Evidence-Based Practice Review Report

Theme: School Based Interventions for Special Educational Needs

How Effective is the Incredible Years® Teacher Classroom Management Programme in Improving Conduct and Prosocial Behaviour in Young Children?

Summary

This literature review aims to evaluate the effectiveness of the Incredible Years® Teacher Classroom Management (IY®-TCM) (Webster-Stratton & Reid, 2002) training programme, at reducing conduct difficulties and increasing prosocial skills in children aged three to eight years identified as being at 'high risk' for social emotional and mental health (SEMH) difficulties. IY®-TCM is a universal preventative intervention, training classroom teachers to build positive relationships with parents, pupils and between the pupils themselves. IY®-TCM aims 'to promote emotional and social competence and to prevent, reduce, and treat aggression and emotional problems in young children' (www.incredibleyears.com).

A systematic literature search identified five studies meeting inclusion criteria. Quality and relevance of these studies was reviewed according to Gough's (2007) Weight of Evidence framework. Effect sizes were calculated for measures of participants' conduct difficulties and prosocial skills and meta-analysis was used to calculate pooled effect sizes.

Results suggest that the IY®-TCM intervention is effective at reducing conduct difficulties and improving the prosocial skills of young children with SEMH difficulties.

Introduction

The Incredible Years® Series (IY®) (Webster-Stratton, 2005), designed by Professor Carolyn Webster-Stratton, consists of a 'pyramid' of interlocking programmes (Figure 1). These target parents, children and teachers and are designed to work jointly to promote emotional, social, and academic competence and prevent, reduce, and treat behavioural and emotional problems in young children (Webster-Stratton, 2005). Components are designed such that they can also be used as standalone interventions e.g. the Incredible Years Teacher Classroom Management training (IY®-TCM), which focusses on teachers' effective classroom management skills for children aged three to eight.

IY®-TCM promotes teachers' use of proactive approaches: coaching skills; effective discipline; collaboration with parents; and the ability to teach social skills, emotional literacy, self-regulation and problem-solving. It emphasizes relationship building between teachers and children, their parents and peers, covering topics related to cognition, behaviour and affect (Figure 2) in six full day group workshops over a six month period, with peer support and mentoring between sessions. Group methods include: discussion; goal setting and problem solving; peer support; consideration of values; skills training on cognition and social learning; DVD vignettes; role-play, rehearsal and video-modelling; and training in setting individual behaviour plans.

Teachers set personal goals to develop self-management, self-reflection and cognitive self-control (Webster-Stratton & Reid, 2002).

The IY® programme, including IY®-TCM, is one of eleven 'Blueprints for Violence Prevention' programmes identified as effective by the US Department of Justice in a review of over 600 programmes (Mihalic et al., 2004).

Figure 1: The IY® Series Pyramid Structure (www.incredibleyears.com)

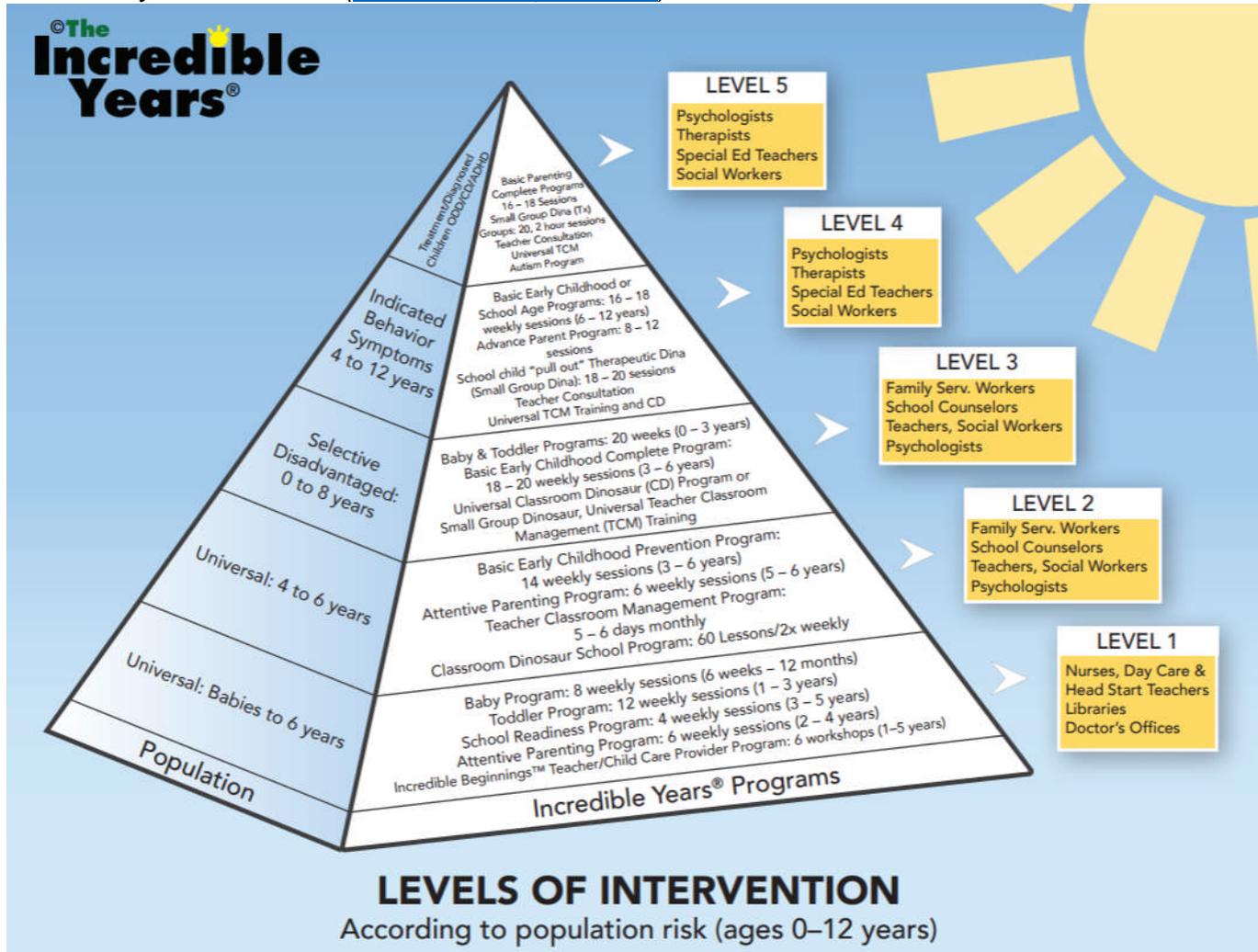
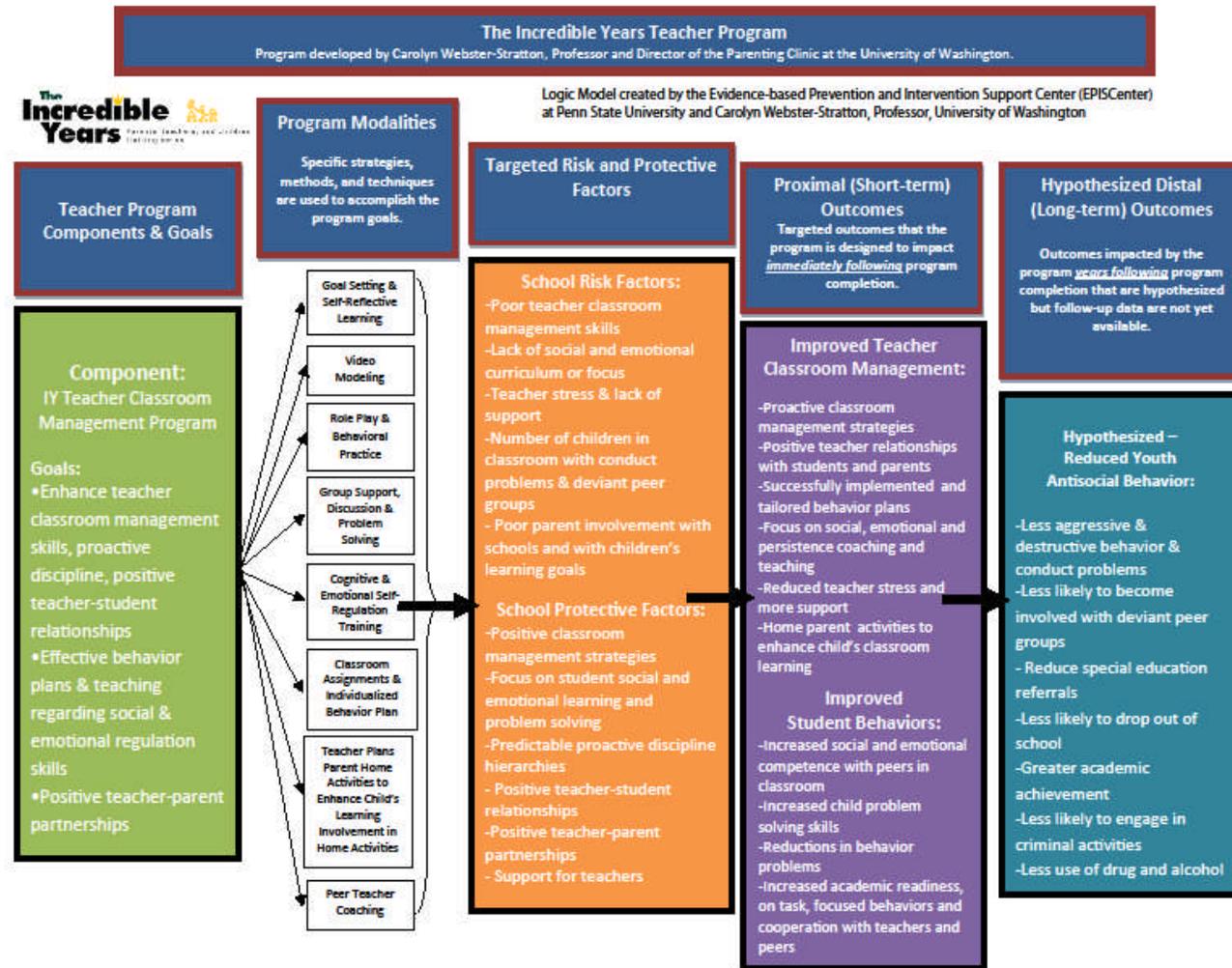


Figure 2: Logic Structure for the IY®-TCM (www.incredibleyears.com)



Psychological Basis

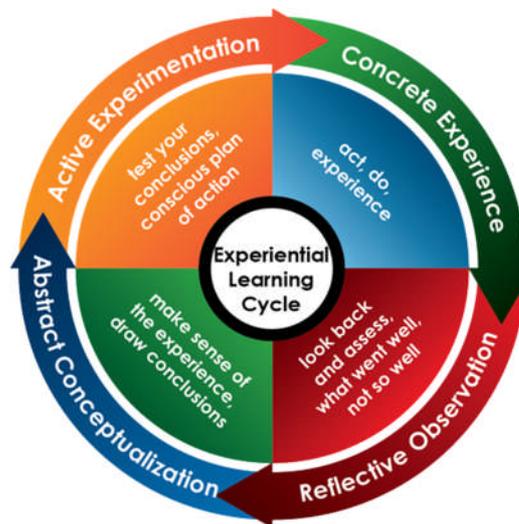
The rationale for IY®-TCM is grounded in cognitive social learning theories on the development of antisocial behaviours in children and young people (CYP). It relies particularly on Patterson's (1976, 1991, 1992) coercion hypothesis that negative reinforcement develops and maintains deviant behaviour in early childhood, cascading into increasingly negative outcomes in adolescence. Bandura's (1969) social learning model recognises that adults serve as both models and reinforcers of behaviour in children. IY®-TCM also draws on Piagetian notions of key developmental stages in affect and cognition (Piaget, 1981), recognising that initial early school experiences can provide the basis for later behaviours and attitudes. Early intervention avoids behaviours becoming entrenched and developing into additional risk factors e.g. academic difficulties, poor attendance and inappropriate peer groups (Dishion & Tipsford, 2011).

In emphasizing the promotion of positive relationships between teachers and their pupils, IY®-TCM draws on Attachment Theory (Bowlby, 1969; Ainsworth, 1989). Teachers' abilities to make positive connections with children and contribute to a positive environment and foundation for learning depend on consistency and sensitivity in their responses. Their ability to manage the classroom environment is critical to children's wellbeing (Reddy et al., 2013) given that teacher-pupil interactions are known to impact on emotional, behavioural and cognitive outcomes (Hamre & Pianta, 2001; Martin & Collie, 2016).

IY®-TCM addresses evidence that poor emotional regulation in early childhood is linked to defiant and aggressive behaviour (Calkins et al. 1998; Schatz et al. 2008). It capitalizes on the role of teachers in supporting the development of children’s emotional regulation skills (Panfile & Laible, 2012; Ahn, 2005), recognising that social competence is a more powerful predictor of long-term academic and career outcomes than cognitive ability (Gutman et al. DfE, 2003; Weare, 2010).

The group training model employed draws on principles of modelling and self-efficacy from Social Learning Theory (Bandura, 1982), and from Kolb’s (1984) Experiential Learning Theory. This cyclical model combines experience, perception, cognition, and behaviour (Figure 3).

Figure 3: Kolb’s (1984) Experiential Learning Cycle



(image from http://elearningwiki.com/index.php?title=Course_structure)

Weare (2010) and Mihalic (2004) emphasise the importance of implementation fidelity given issues with dissemination of evidence-based programmes (e.g. Welch et al., 2006) and the impact this can have on

effectiveness (Gottfredson et al., 2000). IY®-TCM uses evidence on optimal training and mentoring for short and long term implementation (Welch et al., 2006), promoting fidelity in several ways e.g. selection of group leaders and teachers; training workshops with ongoing mentoring; peer support networks; leader accreditation; emphasising infrastructure and support; and ongoing quality assurance/evaluation of fidelity (Webster-Stratton et al., 2011).

Relevance to Educational Psychology Practice

The recent Green Paper on CYP Mental Health Provision (DoH & DoE, 2017) has re-emphasised the importance of addressing this need. In the UK, one in five students is affected by behavioural problems (BMA, 2013 & 2016).

Nearly 8% of children between the ages of five and ten have a diagnosable mental health disorder (Green et al., 2005) and between three and seven percent of children aged five to fifteen meet diagnostic criteria for conduct disorder (NICE, 2013). Early-onset conduct problems are the most common mental health concern in childhood (Moffitt & Scott, 2008).

Long-term, the outcomes for these CYP are very poor (Gutman et al., 2012; Jones et al., 2015; DoH & DoE, 2017), particularly for those with aggressive behaviours, who are at significantly higher risk of involvement in adolescent and adult criminal violence, drug misuse, and poor economic outcomes e.g. unemployment (Snyder, 2001; Fergusson et al., 2005; DoH & DoE, 2017).

They are also at significant risk for academic underachievement, disengagement, poor attendance and poor teacher-student relationships (Baker et al., 2008; Webster-Stratton et al., 2008).

Teachers have been reporting increasing levels of behavioural problems (Hutchings et al., 2011) with classroom management recognised as one of the most important, yet challenging aspects of teaching (Reinke et al., 2011). The development of positive relationships in the learning environment is a core component of evidence-based management in early educational settings (Simonsen et al., 2008). Disruptive behaviour has significant effects on teachers and the classroom environment (Split et al., 2011). Therefore, decreasing the incidence of externalizing problems and increasing teachers' skills in managing classrooms promotes academic achievement for all children by allowing more time to be spent on learning (Sutherland et al., 2008; Werner, 1999). Farrell et al. (2006) highlight the contribution of EPs in supporting teachers working with children with behavioural, emotional and social difficulties e.g. through training and strategic/capacity building work.

Given that it has been established that early intervention is both more effective and more cost-effective (Durlak et al., 2011; Allen, 2011), it is imperative that EPs are able to identify, recommend and support the implementation of such programmes in schools. IY®-TCM addresses much of the British Medical Association's (BMA, 2016) eleven principles for the improvement of health and wellbeing in CYP e.g. reducing inequality, early intervention, evidence-based practice, and support for families and professionals.

The IY® series has grown in popularity and is used in several countries (Figure 4). Given the level of investment and prevalence of use, it is essential to review its effectiveness. Pidano and Allen (2015) reviewed the

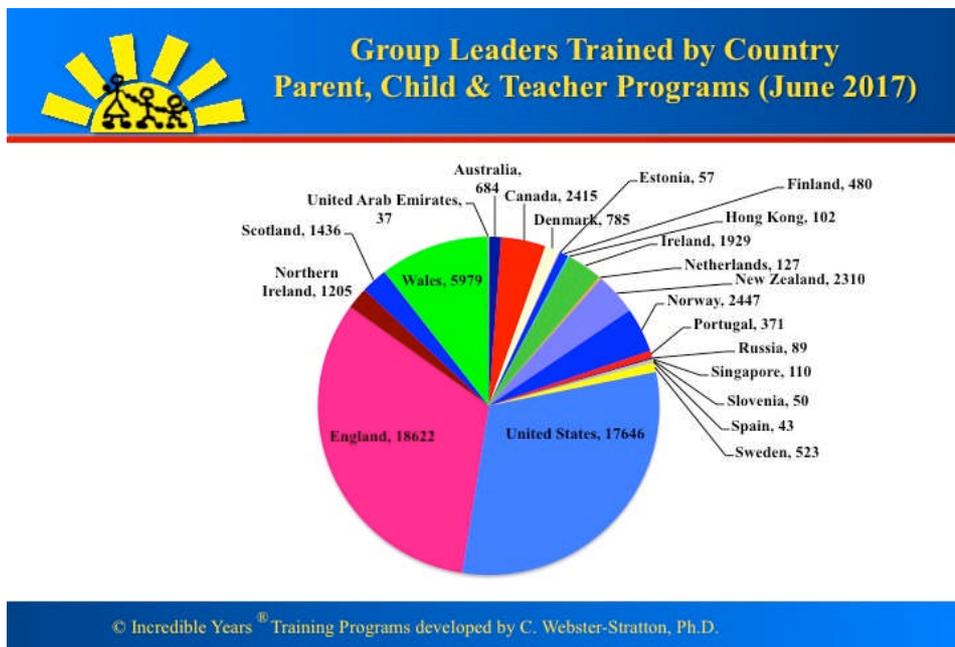
independent evidence for the IY® programmes, but did not specifically consider the effectiveness of IY®-TCM for high risk children. They called for more research evaluating the IY®-TCM programme in particular, and several subsequently published studies are included in the current review.

Review Question

How effective is the IY®-TCM programme in improving conduct and prosocial behaviour in young children with SEMH difficulties?

Figure 4: *IY® Group Leaders Across Countries*

(www.incredibleyears.com/about/incredible-years-series/)



Critical Review of the Evidence Base

Systematic Literature Search

A systematic literature search was conducted between 2nd-16th January 2018, using search terms and the process in Figure 5. Database tools were used to limit and filter search results to peer reviewed journals and English language studies only as per inclusion criteria (Table 1). Five studies met inclusion criteria (Table 2).

Figure 5: *Flow Diagram of the Search/Exclusion Process for Studies*

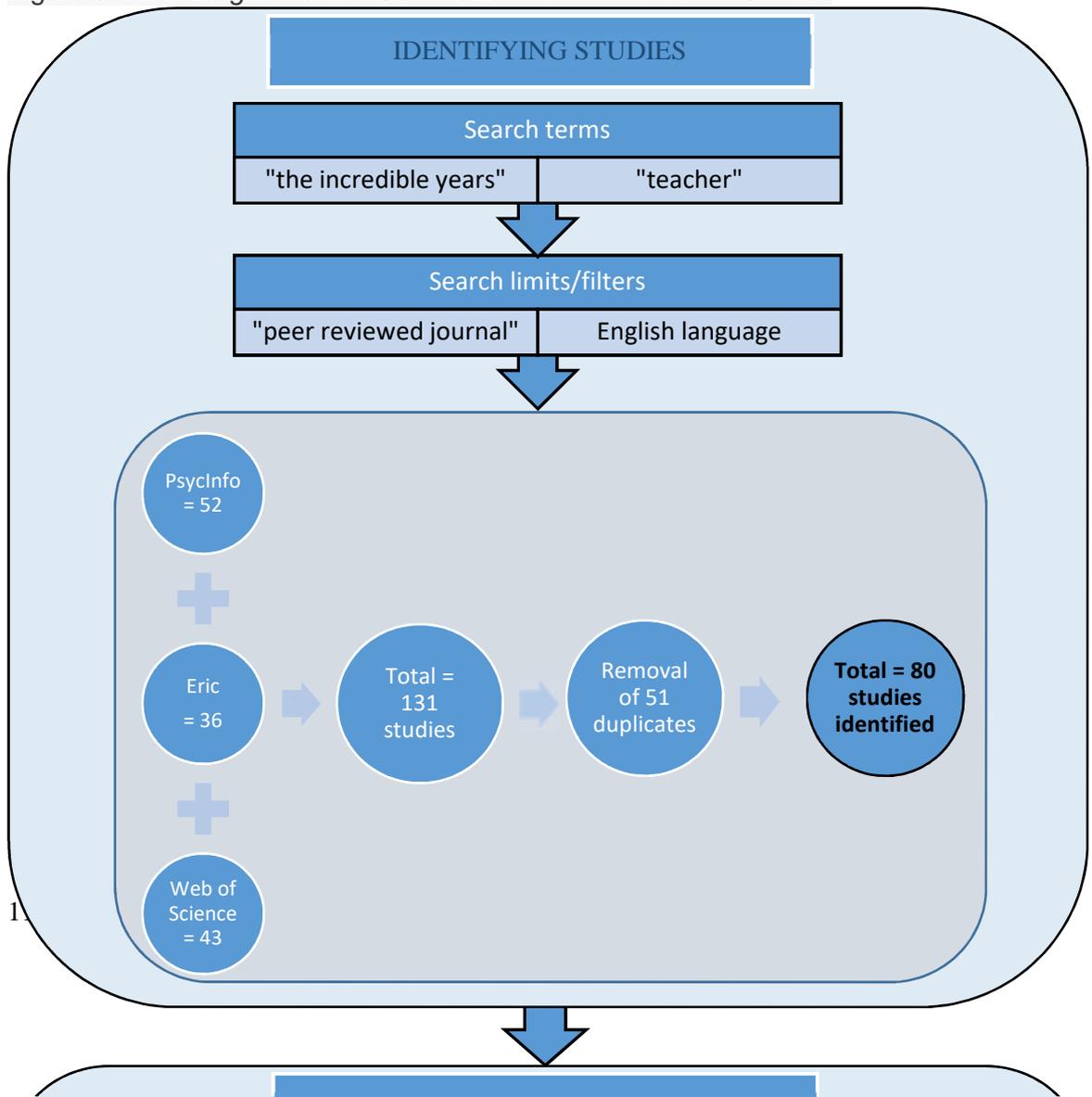


Table 1: *Inclusion and Exclusion Criteria with Rationale*

Study Feature	Inclusion Criteria	Exclusion Criteria	Rationale
1 Intervention	IY@TCM training program only.	Not exclusively IY@TCM e.g. significant adaptations or concurrent interventions.	Review question focusses on IY@-TCM and significant adaptations/concurrent interventions could confound effects.
2 Sample	'High risk' for socioemotional/mental health/behavioural difficulties	Not identified as 'high risk'	Review question focusses on children with behavioural and SEMH difficulties.
3 Publication type	Peer reviewed journal article	Not a peer reviewed journal article	Peer verification prior to publication ensures methodological rigour & quality.
4 Study type	Primary study	Not a primary study e.g. review.	Review question requires completed data.
5 Method	Includes control/comparison group	No control/comparison group	Comparison to assess impact & effect.
6 Outcomes	Measure of children's behaviour/ socioemotional wellbeing.	No measure of child behaviour /socioemotional wellbeing	Review question requires analysis of changes in behaviour.
7 Language	Published in English	Not published in English	To facilitate reading and understanding of the article.

8	Data	Reports data that allows for calculation of effect sizes e.g. pre/post means, standard deviations	Does not report the data necessary for calculating effect size.	To facilitate comparison across studies, effect sizes must be calculated.
9	Authors	Independent researchers/authors	Authors/researchers who benefit directly from the programme	To avoid author bias

Table 2: Full References for the Five Studies Included in this Review.

Full Study Reference
1. 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. <i>British Journal of Psychiatry</i> , 201(2).
2. Fossum, S., Handegard, B. H., & Drugli, M. B. (2017). The Incredible Years Teacher Classroom Management Programme in Kindergartens: Effects of a Universal Preventive Effort. <i>Journal of Child and Family Studies</i> , 26(8).
3. Hickey, G., McGilloway, S., Hyland, L., Leckey, Y., Kelly, P., Bywater, T., Comiskey, C., Lodge, K., Donnelly, M., & O'Neill, D. (2017). Exploring the effects of a universal classroom management training programme on teacher and child behaviour: A group randomised controlled trial and cost analysis. <i>Journal of Early Childhood Research</i> , 15(2).
4. 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. <i>Journal of School Psychology</i> , 51(5).
5. Kirkhaug, B., Drugli, M. B., Handegard, B. H., Lydersen, S., Asheim, M., & Fossum, S. (2016). Does the Incredible Years Teacher Classroom Management Training programme have positive effects for young children exhibiting severe externalizing problems in school?: a quasi-experimental pre-post study. <i>BMC Psychiatry</i> , 16(1).

Quality & Relevance Appraisal

Table 3 outlines an overview of the reviewed studies. These were assessed using Gough's (2007) Weight of Evidence (WoE) approach to critically consider their methodological quality (WoE A), methodological relevance (WoE B) and topic relevance (WoE C) to the review question. An adapted version of Kratochwill's (2003) coding protocol for group-based designs was used to derive WoE A ratings (Appendices B, C & D). WoE B was rated according to a typology as per Petticrew and Roberts (2003) and WoE C according to criteria in Appendix E. WoE D is an average across WoE A, B & C (Table 4).

Table 3: *Mapping the Field*

Study	Sample	Location	Participants	Study Type & Control Group	Outcome Measures (relevant to this study)
1 Baker-Henningham et al. (2012)	225 high risk children (intervention group 113).	Kingston Jamaica	Children (age 3-6yrs) from 24 community preschool classes screened for high levels of teacher-reported conduct problems pre-intervention, based on ICD-10 Diagnostic Criteria for Research. 3 children with highest level in each class evaluated in study.	Cluster RCT (between schools randomisation). Wait list control	Observational measure of child behaviour: Directly observed in-class using researcher-devised observation schedule. Conduct problems teacher-ratings: Sutter-Eyberg Student Behaviour Inventory (SESBI) Conduct problems parent-ratings: Eyberg Child Behaviour Inventory (ECBI) Hyperactivity & attention difficulties teacher-ratings: Conners Global Index Behaviour difficulties and prosocial skills parent and teacher ratings: Strengths and Difficulties Questionnaire (SDQ) Social skills teacher ratings: Preschool and Kindergarten Behavior Scales: Social Skills Scale (PKBS)
2 Fossum et al (2017)	Subgroup of 106 'high risk' children (intervention group 60 'high risk')	Norway	Children aged 3-6yrs from 86 kindergartens screened for aggression on SESBI-R and identified as 'high risk' if $\geq 90^{\text{th}}$ percentile pre-intervention.	Quasi-experimental pre-post design with continuous enrolment (between schools randomisation).	Conduct problems teacher-ratings: Sutter Eyberg Student Behaviour Inventory Total Intensity (SESBI-R) & Child Teacher Report Form (CTRF) – Internalisation, Anger & Aggression scales Prosocial skills & behaviour teacher ratings: Social Competence and

Study	Sample	Location	Participants	Study Type & Control Group	Outcome Measures (relevant to this study)
				Wait list control	Behaviour Evaluation for Teachers (SCBE-T)
3 Hickey et al (2017)	217 children including 63 'high risk' (intervention group 110 children including 33 'high risk')	South-West Ireland	Children (mean age 5.4yrs) from 22 classrooms across 11 schools (7 designated as 'disadvantaged'). Identified as 'high risk' if >12 on pre-intervention teacher SDQ ratings for socioemotional and behavioural difficulties.	Cluster RCT (within schools randomisation). Wait list control	Child behaviour and prosocial skills teacher ratings: Teacher SDQ Observational measure of child behaviour: Teacher Pupil Observation Tool (T-POT)
4 Hutchings et al. (2013)	107 children including 27 'high risk' (intervention group 13 'high risk' children)	North West Wales, UK	Children (ages 3–7 years) from 12 classrooms across 11 primary schools were identified as 'high risk' if above clinical threshold (>15) on pre-intervention teacher SDQ ratings.	Stratified cluster RCT with pair-wise matching of classrooms (between schools randomisation). Wait list control	Observational measure of child behaviour and prosocial skills: Teacher-Pupil Observation Tool (T-POT)
5 Kirkhaug et al (2016)	83 high-risk children (intervention group 45 'high risk' children)	Norway	43 schools = 83 children ages 6-8 assessed by teachers as having severe externalising problems	Quasi-experimental pre-post study with continuous enrolment (between schools randomisation).	Child behaviour teacher ratings: SESBI-R externalising behaviour. Internalising problems & academic performance teacher ratings: Achenbach system of empirically based assessment, teacher report form (ASEBA-TRF)

Study	Sample	Location	Participants	Study Type & Control Group	Outcome Measures (relevant to this study)
				Wait list control	Social skills teacher report: Social Skills Rating System (SSRS)

Table 4: Overview of WoE Ratings:

Author	WoE A: Quality of Methodology	WoE B: Relevance of Methodology	WoE C: Relevance of evidence to the review question	WoE D: Overall weight of evidence
1 Baker-Henningham et al. (2012)	High (2.7)	Medium (2)	Low (1)	Medium (1.9)
2 Fossum et al. (2017)	Medium (2)	Medium (1.5)	Medium (2)	Medium (1.8)
3 Hickey et al. (2017)	Medium (1.7)	Medium (2)	Medium (2)	Medium (1.9)
4 Hutchings et al. (2013)	Low (1)	High (2.5)	High (3)	High (2.2)
5 Kirkhaug et al. (2016)	Medium (1.7)	Medium (1.5)	Medium (2)	Medium (1.7)

Design

Two studies used a quasi-experimental design (Kirkhaug et al., 2016; Fossum et al., 2017) and were given lower WoE B ratings as a result. Two studies used cluster randomised control trials (RCT) with between schools randomisation (Baker-Henningham et al., 2012; Hickey et al., 2017) and were given 'medium' WoE B ratings. These studies used post-hoc tests to establish baseline homogeneity between groups. Hutchings et al., (2013) used a stratified, cluster RCT with pair-wise matching of classrooms and between school randomisation and was given a higher WoE B rating.

All of the included studies used a wait-list comparison group. This is a significant limitation given the Hawthorne effect (Merrett, 2006) whereby changes may arise from involvement in research alone and cannot necessarily be attributed to the intervention itself. It is evident that the extra hours of 'attention' that teachers in the intervention group received during the

IY®-TCM training may have resulted in changes in teacher attitude/behaviour (and therefore in actual or perceived pupil behaviour), irrespective of the curriculum taught during the training sessions. Since this was not controlled for by having at least an 'attention-matched' control group, conclusions about the mechanism for behavioural change must, necessarily, be tenuous.

Participants: Demographics & Recruitment

Across the five studies there were five hundred and four 'high risk' participants aged three to eight years from four different countries. Gender distribution is reported for the whole sample in the studies but not specifically for those with a 'high risk' subset. There were generally more boys than girls reported except for Fossum et al (2017) who had approximately equal numbers and Hickey et al. (2017) who had slightly more girls than boys. Baker-Henningham et al (2012) had the largest sample size and this is reflected in the higher WoE A rating given the need for sufficient power to detect the small to medium range effect sizes expected. The sample in Kirkhaug et al. (2017) was a slightly older demographic (age six to eight) than the other studies (ages three to six or seven), which may impact on effectiveness given the preferential effects of early intervention (Durlak et al., 2011).

Identification and definition of 'high risk' participants across the studies was achieved using a range of different measures. Kirkhaug et al. (2017) and Fossum et al. (2017) used scores above the 90th percentile (the clinical cut-off) for teacher ratings on the Sutter-Eyberg Student Behaviour Inventory (SESBI) Intensity Scale (Eyberg & Pincus, 1999). Hickey et al. (2017) and

Hutchings et al. (2013) used teacher ratings on the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) with the former setting their cut-off for 'high-risk' at twelve (borderline) or above, whereas the latter included only those with scores above 15 ('abnormal') (Goodman, 1997). Participants in Hutchings et al. (2013) may therefore have more marked difficulties than those in Hickey et al. (2017). Given that the focus of the current review is on those with a high degree of need, and Hickey et al.'s (2017) participants do not all reach clinical levels of concern, this is reflected in the lower WoE C rating. Similarly, for Baker-Henningham et al. (2012) who chose the three highest scoring children in each class based on an interviewer-administered teacher-rating scale using ten items from the ICD-10 Diagnostic Criteria for Research for conduct disorder (WHO, 1993), but for whom only just above fifty percent of those classified as 'high risk' scored within the clinical range.

The settings recruited in Fossum et al. (2017) and Kirkhaug et al. (2016) were from rural and urban municipalities in Norway. Kirkhaug et al. (2016) included only participants whose parents spoke Norwegian and report that forty-one percent of the participants in their 'high risk' sample were in receipt of special education assistance in class. The sample in Hickey et al. (2017) was drawn from 8 urban schools, 2 semi-urban schools and 1 rural school, seven of which were designated as 'disadvantaged'. These studies did not provide information on ethnicity of participants. Participants in Hutchings et al. (2012) were all of Caucasian ethnicity and from diverse socioeconomic catchment areas.

The demographics in the Baker-Henningham et al. (2012) study were different given that this took place in a non-OECD country (Jamaica), and in inner-city community-preschool settings in an area of high deprivation. The settings recruited in this study differed from others in that the staff were mostly paraprofessionals. This is reflected in the lower WoE C rating, given that it may impact on outcomes of the intervention, since less experienced staff might benefit more from training than staff with more extensive training at baseline. This may have contributed to the greater number of statistically significant findings in Baker-Henningham et al. (2012). Alternatively, the larger number of statistically significant findings may reflect the larger sample size (and therefore increased power) in this study.

Control for possible confounding variables varied across the studies. Baker-Henningham et al. (2012) controlled for attendance, disability and sibling-pairs. Several of the studies controlled for prior or concurrent exposure to interventions e.g. Kirkhaug et al (2016) and Fossum et al. (2017) excluded any settings that were currently attending or had attended any other evidence based school behaviour intervention program in the previous year. However, these studies recruited their intervention settings from within municipalities that had already implemented IY® Parenting Training – it is unknown if the parents in the sample had received such training. Hutchings et al. (2013) controlled for previous exposure to IY® programs but not for other interventions. Hickey et al. (2017) and Baker-Henningham et al. (2012) do not provide information about exposure to concurrent or previous interventions.

Measures

The studies used a variety of methods for outcome measures, with those using multi-method and multi-source measures achieving higher ratings on these criteria for WoE A. Fossum et al. (2017) and Kirkhaug et al. (2016) had low scores given a lack of triangulation of data due to using teacher ratings only, and Hutchings et al. (2013) for using observation-only. Hickey et al. (2017) achieved a higher rating due to their use of both observation and teacher-ratings, and Baker-Henningham et al. (2012) scored highest due to their combination of teacher, parent and observer ratings.

The measures themselves also varied and the reliability of these contributed towards WoE A. Two studies obtained teacher ratings for conduct using the SDQ (Baker-Henningham et al., 2012; Hickey et al., 2017) and three used the SESBI (Baker-Henningham et al., 2012; Fossum et al., 2017; Kirkhaug et al. 2016). Baker-Henningham et al. (2012) obtained teacher ratings of hyperactivity and attention difficulties using the Conners Global Index (Kollins et al., 2004), and Fossum et al., (2017) obtained measures of internalisation and aggression on the Child Teacher Report Form (CTRF; Achenbach & Rescorla, 2000). Kirkhaug et al (2016) used the Achenbach System of Empirically Based Assessment Teacher Rating Form (ASEBA-TRF; Achenbach & Rescorla, 2001) to measure internalising problems and academic performance.

Teacher ratings of social skills were obtained using the SDQ (Hickey et al., 2017; Baker-Henningham et al., 2012); the Social Skills Rating Scale (SSRS; Gresham & Elliott, 1990) (Kirkhaug et al., 2016); the Social Competence and

Behaviour Evaluation for Teachers (SCBE-T; LaFreniere and Dumas, 1995) (Fossum et al., 2017); or the Preschool and Kindergarten Behaviour Scales-Social Skills Scale (PKBS-SS; Merrell, 1996) (Baker-Henningham et al., 2012). Kirkhaug et al. (2016) also used a population adjusted version of the Student Teacher Relationship Scale (STRS; Pianta, 1996; Drugli et al., 2012) to assess conflict and closeness in teacher-pupil relationships.

Baker-Henningham et al. (2012) obtained parent ratings for conduct and social skills using the SDQ and the Eyberg Child Behaviour Inventory (ECBI; Eyberg, 1978). Hutchings et al. (2013) and Hickey et al. (2017) obtained observational data using the Teacher Pupil Observation Tool (T-POT; Martin et al., 2010) while Baker-Henningham et al. (2012) used a researcher devised protocol.

Low Cronbach's alpha (α) levels were reported for the SDQ ($\alpha=0.56$, Hickey et al., 2017) and the T-POT ($\alpha=0.49$, Hutchings et al., 2013) and this is reflected in the WoE A ratings for these studies.

Implementation fidelity

Given the inbuilt fidelity assurance processes in IY®TCM as described earlier, most of the studies were rated highly for implementation fidelity. However, Fossum et al. (2017) received a lower WoE A rating due to barriers set by the funders of the study who prioritised independence between the research and implementation, thus limiting availability of fidelity data. Hickey et al. (2017) evaluated facilitator reported fidelity but did not validate this independently.

Baker-Henningham et al. (2012) used both facilitator report and teacher evaluations to monitor fidelity. However, given their cultural sensitivity adaptations (e.g. videos set in Jamaican classrooms), and the extra training days (eight days instead of the standard six), this is reflected in their lower WoE C rating.

Effect Sizes and Findings

Each study used different methods for calculating their effect sizes (ES) therefore, in order to facilitate cross-comparison, the ES reported in this review (see Tables 5 & 6) were calculated using between group post-intervention means and standard deviations to calculate Hedge's *g* (Hedges, 1981). All the studies demonstrated homogeneity between groups at pre-test and provided the necessary data for calculation:

$$\text{Hedges' } g = \frac{M_1 - M_2}{SD_{pooled}^*}$$

Only outcome measures of relevance to the review were included (e.g. omission of teacher-level or classroom-level outcomes). ESs can be interpreted according to Cohen's (1988) conventions as small (0.2), medium (0.5) and large (0.8).

Table 5: Hedge's *g* ES and Confidence Intervals (CI) for Behaviour Outcome Measures (* Statistically significant at the 5% probability level)

Authors	WoE D	n	Outcome measure	ES	C.I. (95%)	ES descriptor
Baker-Henningham et al. (2012)	1.9	225	Observer rated conduct problems	0.33	0.07 - 0.60	small*
			Teacher rated conduct problems (SESBI)	0.49	0.22 - 0.75	medium*
			Teacher reported behaviour difficulties (SDQ)	0.27	0.01 - 0.54	small*
			Teacher reported ADHD symptoms (Conners)	0.50	0.24 - 0.77	medium*
			Parent reported behaviour difficulties (ECBI)	0.23	-0.03 - 0.49	small
			Parent reported behaviour difficulties (SDQ)	0.00	-0.26 - 0.26	none
Fossum et al (2017)	1.8	83	Teacher reported behaviour intensity (SESBI)	0.48	0.28 - 0.68	medium*
		82	Teacher reported behaviour problems (SESBI)	0.20	-1.20 - 1.6	small
		83	Teacher reported internalising (C-TRF)	0.24	-1.43 - 1.91	small
		68	Teacher reported aggression (C-TRF)	0.52	0.04 - 1.02	medium*
		66	Teacher reported attention (C-TRF)	0.45	-0.04 - 0.95	medium
Hickey et al (2017)	1.9	63	Teacher reported emotional symptoms (SDQ)	0.30	-0.19 - 0.81	small
			Teacher reported conduct problems (SDQ)	0.23	-0.27 - 0.73	small
			Teacher reported hyperactivity (SDQ)	0.43	-0.07 - 0.94	small
			Teacher reported total difficulties (SDQ)	0.69	0.19 - 1.21	medium*
			Teacher reported impact of difficulties (SDQ)	0.52	0.02 - 1.03	medium*
Hutchings et al. (2013)	2.4	27	Observer rated child compliance (T-POT)	0.32	-0.44 - 1.11	small
			Observer rated child noncompliance (T-POT)	0.57	-0.20 - 1.37	medium
			Observer rated child negative to teacher (T-POT)	0.68	-0.09 - 1.5	medium
			Observer rated child deviance (T-POT)	0.49	-0.28 - 1.28	medium
			Observer rated child off-task behaviour (T-POT)	0.84	0.06 - 1.67	large*
Kirkhaug et al (2016)	1.7	83	Teacher rated child externalising behaviour (SESBI-R)	0.26	-0.22 - 0.76	small
			Teacher rated child-teacher conflict (STRS)	0.18	-0.31 - 0.67	none

Teacher rated child internalising (TRF) **0.39** -0.10 - 0.89 small

Table 6: Hedge's *g* ES and CI for Social Skills Outcome Measures (* Statistically significant at the 5% probability level)

Authors	WoE	D	n	Outcome measure	ES	C.I. (95%)	ES descriptor
Baker-Henningham et al. (2012)	1.9	225		Teacher reported social skills (PKBS)	0.32	0.06 - 0.59	small*
				Teacher reported prosocial behaviour (SDQ)	0.59	0.32 - 0.86	medium*
				Parent reported prosocial skills (SDQ)	0.08	-0.18 - 0.35	none
Fossum et al (2017)	1.8	84		Teacher reported social competence (SCBE-T)	0.29	-0.14 - 0.73	small
Hickey et al (2017)	1.9	63		Teacher reported peer problems (SDQ)	0.79	0.28 - 1.31	large*
				Teacher reported prosocial behaviour (SDQ)	0.54	0.04 - 1.05	medium*
Hutchings et al. (2013)	2.4	27		Observer rated child prosocial behaviour (T-POT)	-0.10	-0.87 - 0.67	none
Kirkhaug et al (2016)	1.7	83		Teacher rated child social skills (SSRS)	0.30	-0.19 - 0.8	small
				Teacher rated child-teacher closeness (STRS)	0.26	-0.23 - 0.76	small

Effectiveness of IY®-TCM for conduct problems:

Overall, the ES reported in Table 5 indicate that data from these studies suggest small to medium effects for reduced behavioural difficulties as rated by teachers and observers. While some of the effects do not reach significance at the 5% probability level, this may be attributable to the studies being underpowered due to insufficient sample size. Nonetheless, three of the studies (Baker-Henningham et al., 2012; Fossum et al., 2017; Hickey et al., 2017) found significant, medium effects in favour of a decrease in the intensity or impact of conduct problems as a result of IY®-TCM.

Data from Hutchings et al. (2013) reveals a significant, large ES in favour of IY®-TCM reducing off-task behaviour as measured by independent observers. Baker- Henningham et al. (2012) data suggests a significant, medium effect on Attention Deficit Hyperactivity Disorder symptoms. However, ES from their data for deviant behaviour did not reach significance.

Kirkhaug et al. (2016) findings did not reach significance and this may have been influenced by the high dropout rate in the intervention group. It is concerning that this study experienced high levels of dropout in the intervention but not the control group. Approximately half of these dropouts were due to student/family withdrawal, and half to teacher withdrawal, although the reasons for this are not explored. It would be important to investigate whether this could indicate poor acceptability for parents, students and teachers, and/or feasibility of the intervention for teachers.

Baker-Henningham et al. (2012) were the only study to explore ratings outside the classroom (through parent report), with their data suggesting that positive effects of IY®-TCM did not generalise beyond school.

Effectiveness of IY®-TCM for prosocial skills:

Two studies (Baker-Henningham et al., 2012; Hickey et al., 2017) show significant medium effects in favour of IY®-TCM improving teacher-rated prosocial skills and Hickey et al. (2017) reveals a significant large effect on teacher-reported peer problems. Two further studies observed non-significant trends for small to medium effect sizes in favour of improved teacher-ratings of prosocial behaviour.

Baker-Henningham (2012) study reported no change in parent-ratings of child behaviour, which may suggest that effects do not generalise beyond school.

Clinical significance

Data on percentage change in clinical significance was only reported by Baker-Henningham et al. (2012), who found a significant between-group difference in the proportion of participants scoring within the clinical range post-intervention for teacher reported conduct problems (SESBI-Intensity scale). Fifty percent of the comparison group scored within the clinically significant range on post-measure, as compared to thirty percent of the intervention group (adjusted OR= 0.31 CI= 0.11-0.92, p=0.02). There was also a significant between group difference in the percentage of participants scoring within the clinical range on both the teacher and parent (ECBI

Intensity scale) measures post-intervention. For the comparison group, 23% of the participants were still above both clinical thresholds, as compared to 9% for the intervention group (adjusted OR=0.24, CI=0.08-0.73, p=0.01).

There was no significant difference between groups based on parent measures only.

Meta-analysis

Meta-analysis was conducted using the MetaEssentials (Suurmond et al., 2017) workbook using a random effects model and WoE D ratings as a moderator. Effect sizes for conduct and social skills were compared across the studies in two separate meta-analyses. Where studies reported more than one measure relevant to each outcome, only one was included to avoid bias from duplicating participant-sets. For the 'conduct' meta-analysis, teacher-reported measures of conduct were used where available, to make the comparison more equitable. For Hutchings et al. (2013) observer ratings of Child Deviance were the most comparable measure available. Baker Henningham et al. (2012) and Fossum et al. (2017) both reported Conduct Problems ratings on the SESBI and of the measures reported by Kirkhaug et al. (2016), SESBI teacher ratings of Intensity of Externalising Behaviour difficulties (as opposed to internalising) were felt to be most comparable. For Hickey et al. (2017), the measure that most closely resembled the nature of those in the other studies was teacher SDQ ratings of conduct problems.

For the 'social skills' meta-analysis, Baker-Henningham et al (2012) were the only study with multiple relevant teacher ratings. Their SDQ prosocial ratings were chosen over the PKBS social skills ratings given that the former could

be directly compared with the SDQ prosocial ratings from Hickey et al. (2017).

Funnel plots were examined to check for publication bias (see Appendix F), alongside Egger's test of asymmetry. These supported a low risk of publication bias (conduct: intercept =-0.73, t=-0.87, p=0.45; prosocial: intercept =-2.48, t=-2.78, p=0.07). There were no studies outside of the 95% confidence intervals and trim and fill adjustments estimated no missing studies. Low risk of heterogeneity was found (conduct: Q=1.42, I^2 =0.00; prosocial: Q=4.19, I^2 =4.62%).

Figure 6: *Conduct Meta-Analysis*

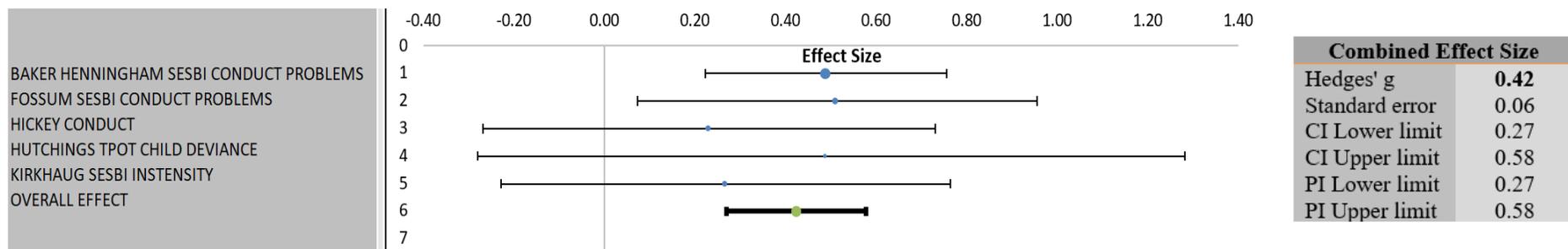
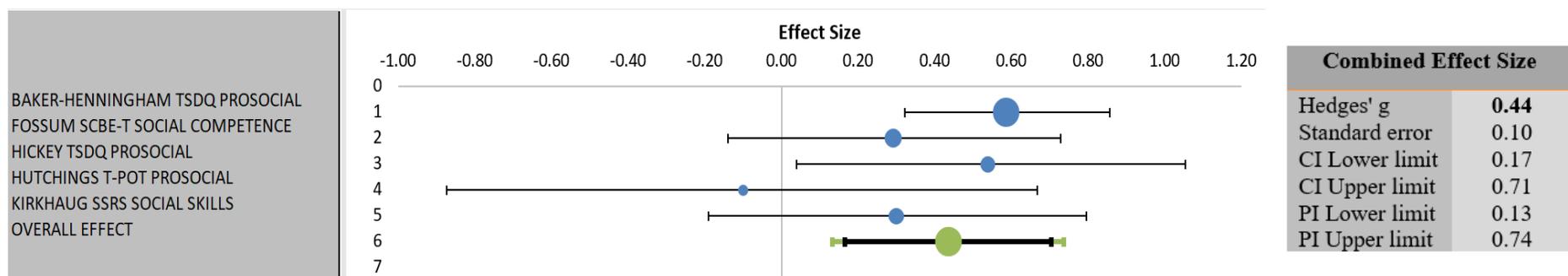


Figure 7: *Social Skills Meta-Analysis*



Conclusions & Recommendations

The present meta-analysis reviews data from five studies and five hundred and four participants aged three to eight years across four different countries. Results support the evidence-base for the effectiveness of the universal preventative IY®-TCM intervention in improving conduct difficulties and promoting prosocial behaviour in the classroom, for young children with SEMH difficulties. Results show little evidence of publication bias or heterogeneity in the sample and the small to medium pooled ESs suggest that it may be appropriate for EPs to consider recommending IY®-TCM as a universal preventative intervention to settings with concerns about young children with significant SEMH needs.

Limitations

There was a large variation in sample sizes, with three out of the five studies underpowered for small-medium effect sizes (Hickey et al., 2017; Hutchings et al., 2013; Kirkhaug et al., 2016). This may explain some of the non-significant findings in these studies. Further RCTs should ensure sufficient power through recruitment of an appropriate number of high-risk participants.

Two of the studies relied on teacher-ratings only (Fossum et al 2017, Kirkhaug, 2016). Given that teachers were the agents of change in the intervention, and were not blind to the randomisation, triangulation of data through blind observer ratings should be used to avoid teacher-bias in outcome measures.

While the studies all included control groups, these were all 'business-as-usual' controls and for the purposes of validity, future research would benefit from attention-controls to establish that the mechanism for behaviour modification cannot be attributed to the Hawthorne effect (Merrell, 2006).

There is a need for further research to verify whether effects of the IY®-TCM generalise beyond the school, given that the parent rating data obtained by Baker-Henningham et al. (2012) indicated no change in behaviour at home.

The higher level of attrition in the intervention group in the Kirkhaug et al. (2017) study should be explored to assess whether it indicates low acceptability/feasibility of the intervention. However, teacher acceptability was not reported to present an issue in the four other studies reviewed and Hutchings et al. (2007) reported teacher satisfaction with the programme. Hansford et al. (2015) are currently researching influences on teacher uptake.

References & Appendices

References

- Achenbach, T. M., & Rescorla, L. A. (2000). *Manual for the ASEBA preschool forms & profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.
- Achenbach, T.M. & Rescorla, L.A. (2001) *Manual for the ASEBA school-age forms & profiles*. Burlington: University of Vermont, Research Center for Children, Youth, & Families.
- Ahn, H.J. (2005). Child care teachers' beliefs and practices regarding socialization of emotion in young children. *Journal of Early Childhood Teacher Education*, 26, 283-295
- Ainsworth, M. (1989) Attachment beyond infancy. *American Psychologist*, 44, pp. 709-716
- Allen, G. (2011). *Early intervention: The next steps—An independent report to Her Majesty's Government*. London, England: HM Government.
- Baker, J. A., Grant, S., & Morlock, L. (2008). The teacher-student relationship as a developmental context for children with internalizing or externalizing behaviour problems. *School Psychology Quarterly*, 23(1), 3-15.
- 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. *British Journal of Psychiatry*, 201(2), 101–108.
<https://doi.org/10.1192/bjp.bp.111.096834>
- Bandura, A. (1969) Social learning of moral judgments. *Journal of Personality and Social Psychology*, 11 (3), 275-279.
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*. 37 (2): 122–147
- Bowlby, J. (1969) *Attachment and Loss*. in <https://www.psychologistworld.com/developmental/attachment-theory#references> (accessed January 2018)
- British Medical Association (BMA) (2013). *Growing up in the UK: Ensuring a healthy future for our children*. <https://www.bma.org.uk/collective-voice/policy-and-research/public-and-population-health/child-health/growing-up-in-the-uk> (accessed January 2018)

- British Medical Association (BMA) (2016). *Growing up in the UK 2016: progress report* <https://www.bma.org.uk/collective-voice/policy-and-research/public-and-population-health/child-health/growing-up-in-the-uk-review> (accessed January 2018)
- Calkins, S.D., Smith, C.L., Gill, K.L. & Johnson, M.C. (1998). Maternal interactive styles across contexts: Relations to emotional, behavioral and physiological regulation during toddlerhood. *Social Development*, 7, 350-369
- Cohen J. (1988) *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Hillsdale, New Jersey: Erlbaum.
- Department of Health, Department for Education. (2017) Transforming Children and Young People's Mental Health Provision: A Green Paper. www.gov.uk/government/consultations/transforming-children-and-young-peoples-mental-health-provision-a-green-paper. (Accessed January 2018)
- DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled clinical trials*, 7(3), 177-188.
- Dishion, T. J., & Tipsord, J. M. (2011). Peer Contagion in Child and Adolescent Social and Emotional Development. *Annual Review of Psychology*, 62, 189–214.
- Dodge, K.A., Greenberg, M.T., Malone, P.S. (2008) Testing an idealized dynamic cascade model of the development of serious violence in adolescence. *Child Development*, 79(6):1907–27.
- Drugli, M.B. & Hjemdal, O. (2012) Factor structure of the student–teacher relationship scale for Norwegian school-age children explored with confirmatory factor analysis. *Scandinavian Journal of Education Resources*, 57(5):457–66.
- Durlak, J.A., Weissberg, R.P., Dymnicki, A.B., Taylor, R.D., & Schellinger, K.B. (2011) The impact of enhancing students' social and emotional learning: a meta-analysis of school-based universal interventions. *Child Development*, 82(1):405–32.
- Eyberg, S., & Pincus, D. (1999). *Eyberg Child Behavior Inventory & Sutter-Eyberg Student Behavior Inventory-Revised: Professional Manual*. Odessa, FL: Psychological Assessment Resources.
- Eyberg, S.M. & Ross, A.W. (1978) Assessment of child behaviour problems: the validation of a new inventory. *Journal of Clinical Child & Adolescent Psychology*, 7: 113–6.
- Farrell, P., Woods, K., Lewis, S., Rooney, S., Squires, G. & O' Connor, M. (2006). *A review of the functions and contribution of educational psychologists in England and Wales in light of "Every Child Matters: Change for Children"*. London: DfES

- Fergusson, D.M., Horwood, L.J., & Ridder, E.M. (2005) Show me the child at seven: the consequences of conduct problems in childhood for psychosocial functioning in adulthood. *J Child Psychol Psychiatry* 2005; 46: 837–49.
- Fossum, S., Handegard, B. H., & Drugli, M. B. (2017). The Incredible Years Teacher Classroom Management Program 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>
- Goodman, R. (1997) The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, 38, 581-586.
- Gottfredson, G. D., Gottfredson, D. C., Czeh, E. R., Cantor, D., Crosse, S. B., & Hantman, I. (2000). *National Study of Delinquency Prevention in Schools. Final Report*. <https://www.ncjrs.gov/pdffiles1/nij/grants/194116.pdf> (retrieved January 2018)
- Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Applied and Practice-Based Research. Special Edition of Research Papers in Education*, 22(2), 213–228. <http://doi.org/10.1080/02671520701296189>
- Green, H., McGinnity, A., Meltzer, H., Ford, T., & Goodman, R. (2005) *Mental health of children and young people in Great Britain, 2004*. Basingstoke: Palgrave MacMillan. <http://digital.nhs.uk/catalogue/PUB06116> (accessed January 2018)
- Gresham, F.M. & Elliott, S.N. (1990) *Social skills rating system manual: circle pines*. MN: American Guidance Service.
- Gutman, L.M., Morrison, L., Sameroff, A.J., Cole, R. (2003) Academic growth curve trajectories from 1st grade to 12th grade: Effects of multiple social risk factors and preschool child factors. *Developmental Psychology*, 39(4): 777-790
- Gutman, L.M., & Vorhaus, J., Department for Education (DFE), corp creator. (2012) *The impact of pupil behaviour and wellbeing on educational outcomes*. [Research report DFE RR253]
- Hamre, B.K. & Pianta, R.C. (2001) Early teacher- student relationships and the trajectory of children's school outcomes through eighth grad. *Child Development*, 2:625–38
- Hattie J. (2009) *Visible learning: a synthesis of over 800 meta-analyses relating to achievement*. London: Routledge in Kirkhaug, B., Drugli, M. B., Handegard, B. H., Lydersen, S., Asheim, M., & Fossum, S. (2016). Does the Incredible Years Teacher Classroom Management Training program have positive effects for young children exhibiting severe

externalizing problems in school?: a quasi-experimental pre-post study. *BMC Psychiatry*, 16.

Hickey, G., McGilloway, S., Hyland, L., Leckey, Y., Kelly, P., Bywater, T., O'Neill, D. (2017). Exploring the effects of a universal classroom management training program on teacher and child behaviour: A group randomised controlled trial and cost analysis. *Journal of Early Childhood Research*, 15(2), 174–194.
<https://doi.org/10.1177/1476718X15579747>

Hughes, K., Bellis, M., Hardcastle, K., Sethi, D., Butchart, A., Mikton, C., Jones, L., & Dunne, M. (2017) The effect of multiple adverse childhood experiences on health: a systematic review and Transforming children and young people's mental health provision meta-analysis. *Lancet Public Health*, 2.
[http://www.thelancet.com/pdfs/journals/lanpub/PIIS2468-2667\(17\)30118-4.pdf](http://www.thelancet.com/pdfs/journals/lanpub/PIIS2468-2667(17)30118-4.pdf) (accessed January 2018)

Hutchings, J., Daley, D., Jones, K., Martin, P., Bywater, T., & Gwynn, R. (2007). Early results from developing and researching the Webster-Stratton Incredible Years teacher classroom management training program in North West Wales. *Journal of Children's Services*, 2, 15-26.

Hutchings, J., Williams, M. E., Martin, P., & Pritchard, R. O. (2011). Introduction of the Incredible Years Child and Teacher programs in Gwynedd: Levels of behavioral difficulties among young welsh school children. *Welsh Journal of Education*, 15(1), 103-115.

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>

Jones, D. E., Greenberg, M., & Crowley, M. (2015). Early social-emotional functioning and public health: The relationship between kindergarten social competence and future wellness. *American Journal of Public Health*, 105(11), 2283-2290.

Kirkhaug, B., Drugli, M. B., Handegard, B. H., Lydersen, S., Asheim, M., & Fossum, S. (2016). Does the Incredible Years Teacher Classroom Management Training program have positive effects for young children exhibiting severe externalizing problems in school?: a quasi-experimental pre-post study. *BMC Psychiatry*, 16.
<https://doi.org/10.1186/s12888-016-1077-1>

Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development* (Vol. 1). Englewood Cliffs, NJ: Prentice-Hall.

Kollins, S.H., Epstein, J.N., Conners, K.C. (2004) *Conners' Rating Scales – Revised*. In *The Use of Psychological Testing for Treatment Planning*

and Outcomes Assessment (3rd edn) (ed. ME Maruish): 215–32.
Lawrence Erlbaum Associates.

- Kratochwill, T.R., (Chair) (2003) Task Force on Evidence-Based Interventions in School Psychology. *Procedural and Coding Manual for Group-Based Intervention Research Design*. Retrieved from <http://www.indiana.edu/~ebi/documents/workingfiles/EBImanual1.pdf> retrieved January 2018.
- Martin, A. & Collie, R. (2016). The role of teacher-student relationships in unlocking students' academic potential: Exploring motivation, engagement, resilience, adaptability, goals, and instruction. In Wentzel, K.R. & Ramani, G. (Eds) *Handbook of social influences on social-emotional, motivation, and cognitive outcomes in school contexts*. Routledge.
- Martin, P.A., Daley, D., Hutchings, J., et al. (2010) The Teacher–Pupil Observation Tool (T-POT): development and testing of a new classroom observation measure. *School Psychology International* 31(3): 229–249.
- Merrell, K.W. (1996) Assessment of social skills and behaviour problems in early childhood: the Preschool and Kindergarten Behavior Scales. *Journal of Early Intervention*, 20: 132–45.
- Merrett, F. (2006). Reflections on the Hawthorne effect. *Educational Psychology*, 26(1), 143-146.
- Mihalic, S. (2004). The importance of implementation fidelity. *Emotional and Behavioral Disorders in Youth*, 4(4), 83-105.
- Mihalic, S., Fagan, A., Irwin, K., Ballard, D., & Elliott, D. (2004) *Blueprints for Violence Prevention*. Washington, D.C.: Office of Juvenile Justice and Delinquency Prevention, Office of Justice Programs, U.S. Department of Justice. <https://www.ncjrs.gov/pdffiles1/ojjdp/204274.pdf> (retrieved February 2018)
- Moffitt, T., & Scott, S. (2008) Conduct disorders in childhood and adolescence. in 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. *British Journal of Psychiatry*, 201(2), 101–108.
- National Institute for Health and Care Excellence (NICE) (2013) *Antisocial behaviour and conduct disorders in children and young people: recognition and management Clinical guideline [CG158]* <https://www.nice.org.uk/guidance/cg158/evidence/conduct-disorders-in-children-and-young-people-full-guideline-189848413> (retrieved January 2018)
- Petticrew, M., & Roberts, H. (2003). Evidence, hierarchies, and typologies: Horses for courses. *Journal of Epidemiology and Community Health*, 57(7), 527–529. <https://doi.org/10.1136/jech.57.7.527>

- 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/http://dx.doi.org/10.1007/s10826-014-9991-7>
- Schatz, J.N., Smith, L.E., Borkowski, J.G., Whitman, T.L. & Keogh, D.A. (2008) Maltreatment risk, self-regulation, and maladjustment in at-risk children. *Child Abuse and Neglect*, 32, 972-982.
- Simonsen, B., Fairbanks, S., Briesch, A., Myers, D., & Sugai, G. (2008). Evidence-based practices in classroom management: Considerations for research to practice. *Education and treatment of children*, 31(3), 351-380.
- Snyder, H. (2001). Child delinquents. Thousand Oaks, CA: Sage. In Webster-Stratton, C., & Reid, J. (2002). The Incredible Years classroom management teacher training program: Content, methods, and process. *Unpublished manuscript, Incredible Years*, University of Washington, Seattle. <http://www.incredibleyears.com/wp-content/uploads/teacher-training-program-classroom-mgmt-02.pdf> retrieved January 2018.
- Split, J. L., Koomen, H. M. Y., & Thijs, J. T. (2011). Teacher Wellbeing: The Importance of Teacher--Student Relationships. *Educational Psychology Review*, 23(4), 457–477. <https://doi.org/10.1007/s10648-011-9170-y>
- Sutherland, K. S., Lewis-Palmer, T., Stichter, J., & Morgan, P. L. (2008). Examining the influence of teacher behaviour and classroom context on the behavioral and academic outcomes for students with emotional or behavioral disorders. *The Journal of Special Education*, 41(4), 223-233.
- Weare, K. (2010). What works in promoting social and emotional well-being and responding to mental health problems in schools ? Advice for Schools and Framework Document. *Advances in School Mental Health Promotion*, 3(1), 5–17. Retrieved from http://www.ncb.org.uk/media/1197143/ncb_framework_for_promoting_well-being_and_responding_to_mental_health_in_schools.pdf
- Webster-Stratton, C., & Reid, J. (2002). The Incredible Years classroom management teacher training program: Content, methods, and process. *Unpublished manuscript, Incredible Years*, University of Washington, Seattle. <http://www.incredibleyears.com/wp-content/uploads/teacher-training-program-classroom-mgmt-02.pdf> Retrieved January 2018
- Webster-Stratton, C., Reid, M.J., & Hammond, M. (2004) Treating children with early-onset conduct problems: intervention outcomes for parent, child, and teacher training. *Journal of Clinical Child and Adolescent Psychology*, 33(1):105-24.

- Webster-Stratton, C. (2005). The Incredible Years: A training series for the prevention and treatment of conduct problems in young children. In E. D. Hibbs & P. S. Jensen (Eds.), *Psychosocial treatments for child and adolescent disorders: Empirically based strategies for clinical practice* (2nd ed., pp. 507- 555). Washington, DC: American Psychological Association.
- Webster-Stratton, C., Reid, M. J., & Stoolmiller, M. (2008). Preventing conduct problems and improving school readiness: Evaluation of the Incredible Years Teacher and Child training programs in high-risk schools. *Journal of Child Psychology and Psychiatry*, 49, 471-488. doi:10.1111/j.1469- 7610.2007.01861.x
- Webster-Stratton, C., Reinke, W. M., Herman, K. C., & Newcomer, L. L. (2011). The incredible years teacher classroom management training: the methods and principles that support fidelity of training delivery. *School Psychology Review*, 40(4), 509.
- Welch, M., Riley, B., Montgomery, P., Linda, v. T., & Mansi, O. (2006). Implementation research: A synthesis of the literature. *Canadian Journal of Public Health*, 97(4), 315.
- Werner, E. E. (1999). How children become resilient: Observations and cautions. In N. Henderson, B. Benard, & N. Sharp-Light (Eds.), *Resiliency in action: Practical ideas for overcoming risks and building strengths in youth, families, and communities* (pp. 115-134). Gorham, ME: Resiliency in Action.
- World Health Organization. (1993). *The ICD-10 classification of mental and behavioural disorders: diagnostic criteria for research* (Vol. 2). World Health Organization.

Appendices

Appendix A – Excluded Studies & Rationale

Authors	Study citation	Rationale for exclusion
Ford et al. (2012)	Supporting teachers and children in schools: the effectiveness and cost-effectiveness of the incredible years teacher classroom management program in primary school children: a cluster randomised controlled trial, with parallel economic and process evaluations. <i>BMC Public Health</i> , 12.	Study protocol only (exclusion criteria 4)
Rimestad et al (2017)	Combining parent and teacher training for early adhd: A randomized study of effectiveness. <i>Journal of Child and Family Studies</i> ,	Used an author abridged version of the IY®-TCM programme (exclusion criteria 1)
Pidano et al. (2015)	The Incredible Years series: A review of the independent research base. <i>Journal of Child and Family Studies</i> , 24(7), 1898–1916.	Not a primary study (exclusion criteria 4)
Snyder et al. (2011)	The impact of brief teacher training on classroom management and child behaviour in at-risk preschool settings: Mediators and treatment utility. <i>Journal of Applied Developmental Psychology</i> , 32(6), 336–345.	Used an author abridged version of the IY®-TCM (exclusion criteria 1)
Hutchings et al. (2017)	Taking the Incredible Years Child and Teacher Programs to Scale in Wales. <i>Childhood Education</i> , 93(1), 20–28.	Not a primary study (exclusion criteria 4)
Hutchings et al. (2007)	Early results from developing and researching the Webster-Stratton Incredible Years Teacher Classroom Management Training Program in North West Wales. <i>Journal of Children’s Services</i> , 2(3), 15–26	Did not include baseline measures (exclusion criteria 8)
Raver et al. (2008)	Improving preschool classroom processes: Preliminary findings from a randomized trial implemented in Head Start settings. <i>Early Childhood Research Quarterly</i> , 23(1), 10–26.	Used an author adapted version of the IY®-TCM including MH professional

intervention
(exclusion criteria
1)

Appendix B – Coding Protocols for Weight of Evidence A

[Adapted from Task Force on Evidence-Based Interventions in School Psychology, American Psychology Association, Kratochwill, T.R. (2003)]

Coding Protocol

Name of Coder: _____ Anon _____
Date: _____ 16.01.18 _____

Full Study Reference in proper format:

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. *British Journal of Psychiatry*, 201(2), 101–108. <https://doi.org/10.1192/bjp.bp.111.096834>

3= Strong Evidence 2=Promising Evidence 1=Weak Evidence
0=No Evidence

1. General Characteristics

A. General Design Characteristics

A1. Random assignment designs (if random assignment design, select one of the following)

- Completely randomized design
- Randomized block design (between participants, e.g., matched classrooms)
- Randomized block design (within participants)
- Randomized hierarchical design (nested treatments)

A2. Nonrandomized designs (if non-random assignment design, select one of the following)

- Nonrandomized design
- Nonrandomized block design (between participants)
- Nonrandomized block design (within participants)
- Nonrandomized hierarchical design
- Optional coding for Quasi-experimental designs

A3. Overall confidence of judgment on how participants were assigned (select one of the following)

- Very low (little basis)
- Low (guess)
- Moderate (weak inference)
- High (strong inference)
- Very high (explicitly stated)
- N/A
- Unknown/unable to code

B. Appropriate Statistical Analysis

Analysis 1 _____ multilevel multiple regression
L1: children (outcomes), L2: classes, L3: preschool
(intervention)
analyses conducted on intention-to-intervene basis,
assuming no change.

- B1 Appropriate unit of analysis
- B2 Familywise/experimenter wise error rate controlled when applicable
- B3 Sufficiently large N

B4. Total size of sample (start of the study): 225

B5. Intervention group sample size: 113

B6. Control group sample size: 112

Overall rating for Statistical analysis: 3

E. Concurrent or Historical Intervention Exposure

- Current exposure
- Prior exposure
- Unknown

2. Key Features for Coding Studies and Rating Level of Evidence/Support

(Rating Scale: 3= Strong Evidence, 2=Promising Evidence, 1=Weak Evidence, 0=No Evidence)

A. Measurement (Estimating the quality of the measures used to establish effects)

A1 The use of the outcome measures produce reliable scores for the majority of the primary outcomes

- Yes
- No
- Unknown/unable to code

A2 Multi-method (at least two assessment methods used)

- Yes
- No
- N/A
- Unknown/unable to code

A3 Multi-source (at least two sources used self-reports, teachers etc.)

- Yes
- No
- N/A
- Unknown/unable to code

A4 Validity of measures reported (well-known or standardized or norm-referenced are considered good, consider any cultural considerations)

- Yes validated with specific target group
- In part, validated for general population only
- No
- Unknown/unable to code

Overall Rating for measurement: 3

B. Comparison Group

B1 Type of Comparison Group (Select one of the following)

- Typical intervention (typical intervention for that setting, without additions that make up the intervention being evaluated)
- Attention placebo
- Intervention element placebo
- Alternative intervention
- Pharmacotherapy
- No intervention
- Wait list/delayed intervention
- Minimal contact
- Unable to identify type of comparison

B2 Overall confidence of judgment on type of comparison group

- Very low (little basis)
- Low (guess)
- Moderate (weak inference)
- High (strong inference)
- Very high (explicitly stated)
- Unable to identify comparison group

B3 Counterbalancing of change agent (participants who receive intervention from a single therapist/teacher etc were counter-balanced across intervention)

- By change agent
- Statistical (analyse includes a test for intervention)
- Other
- Not reported/None

B4 Group equivalence established (select one of the following)

- Random assignment
- Posthoc matched set
- Statistical matching
- Post hoc test for group equivalence

B5 Equivalent mortality

- Low attrition (less than 20 % for post)
- Low attrition (less than 30% for follow-up)
- Intent to intervene analysis carried out?
Findings_____

Overall rating for Comparison group: 2

D: Educational/clinical significance

Outcome Variables	Pretest	Posttest
D1. Categorical Diagnosis Data	Diagnostic information regarding inclusion into the study presented? YES	Positive change in diagnostic criteria from pre to posttest? YES
D2. Outcome assessed via continuous variables		Positive change in percentage of participants showing clinical improvement from pre to posttest?

		YES
D3. Subjective Evaluation: The importance of behaviour change is evaluated by individuals in direct contact with the participant	Importance of behaviour change is evaluated? YES	Importance of behaviour change from pre to posttest is evaluated positively by individuals in direct contact with the participant? YES
D4. Social Comparison: Behaviour of participant at pre and post is compared to normative data (e.g. a typical peer)	Participants behaviour is compared to normative peer? YES	Participants behaviour has improved from pre to posttest when compared to normative data? YES

Overall rating for Educational/Clinical significance: 3

F. Implementation Fidelity

F1 Evidence of Acceptable Adherence

- Ongoing supervision/consultation
- Coding intervention sessions/lessons or procedures
- Audio/video tape implementation
 - Entire intervention
 - Part of intervention

F2. Manualization (select all that apply)

- Written material involving a detailed account of the exact procedures and the sequence in which they are to be used
- Formal training session that includes a detailed account of the exact procedures and the sequence in which they are to be used
- Written material involving an overview of broad principles and a description of the intervention phases
- Formal or informal training session involving an overview of broad principles and a description of the intervention phases

F3. Adaptation procedures are specified (select one):

- Yes
- No
- Unknown/unable to code

Overall rating for Implementation Fidelity: 3

Summary of Evidence

Indicator	Overall evidence rating: 0-3	Description of evidence: Strong Promising Weak No/limited evidence Or Descriptive ratings
General Characteristics		
Design		Cluster RCT
Statistical Treatment	3	Strong
Concurrent/ historical intervention exposure		Unknown
Key features		
Measurement	3	Strong
Comparison group	2	Promising
Educational/clinical significance	3	Strong
Implementation fidelity	3	Strong
Overall WoE A	2.8	Strong

--	--	--

Appendix C - Modifications to Kratochwill Coding Protocol & Rationale

Item/section removed	Rationale for exclusion from coding protocol
Domain	All are school-based intervention as per inclusion criteria
Name of intervention	All are Incredible Years Teacher Classroom Management as per inclusion criteria
Study ID number	N/A
Type of publication	All peer-reviewed journal articles as per inclusion criteria
Section 1: B7 & B8	Not relevant to quantitative analysis
Section 1: C & D	All use the same programme as per inclusion criteria
Section 2: C	Statistical significance of outcomes will be calculated separately
Section 2: G	Not replication studies
Section 2: H	All intervention implemented in schools as per inclusion criteria
Section 2: I	No follow-up assessments
Section 3	Descriptive/supplemental criteria irrelevant or already addressed in mapping the field.

Appendix D – Breakdown of Coding Protocol Scores Across Studies

Author	Statistical Treatment	Measurement	Comparison Group	Educational/Clinical Significance	Implementation Fidelity	Overall WoE A
1 Baker-Henningham et al. (2012)	3	3	2	3	3	2.8
2 Fossum et al (2017)	3	1	2	2	1	1.8
3 Hickey et al (2017)	2	1	2	2	2	1.8
4 Hutchings et al. (2013)	1	0	2	1	3	1.4
5 Kirkhaug et al (2016)	2	2	1	1	3	1.8

Appendix E –Weight of Evidence B & C Rationale

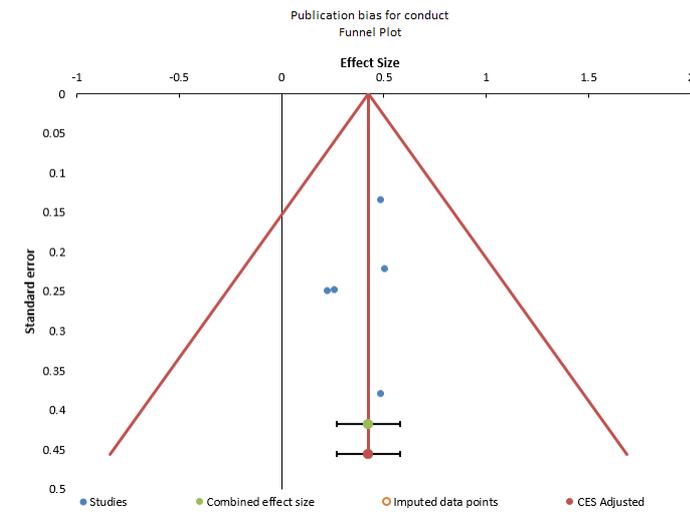
Rationale for WoE B ratings:

Criteria	Descriptor & Score
Full randomised control trial	High (3)
Cluster RCT with pairwise matching	Medium-High (2.5)
Cluster RCT with post-hoc matching	Medium (2)
Quasi-experimental design	Medium-low (1.5)
Small n/case design	Low (1)

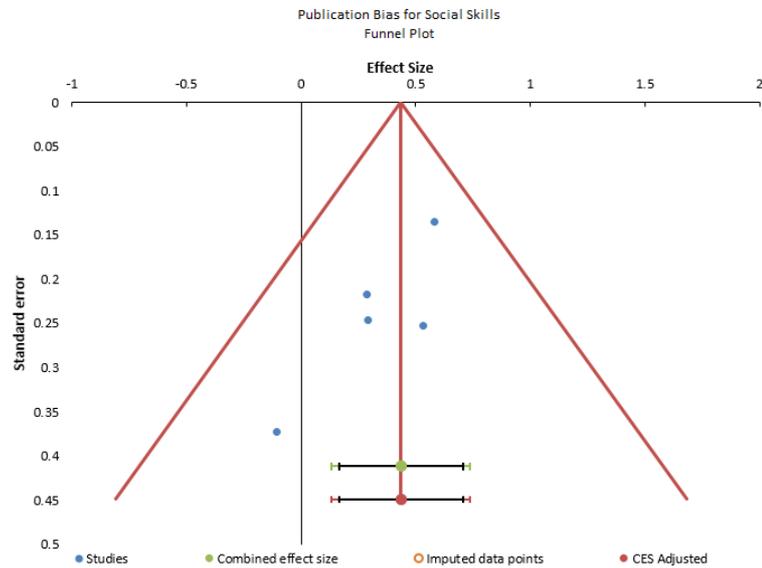
Rationale for WoE C ratings:

Intervention	Participants	Location	Descriptor & Score
Full training delivered	Qualified teachers. Children with clinical levels of concern	UK	High (3)
Minor adaptations to intervention (<+/- 2 days)	Qualified teachers. Children with elevated scores.	OECD countries with educational systems that have broad similarities to those in the UK	Medium (2)
Significant adaptations to length/content of training (>+/- 2 days)	No teacher credentials/not fully qualified. No measurable criteria for children identified as 'high risk'.	Schools in non-OECD countries in which the educational systems are significantly different to those in UK	Low (1)

Appendix F – Meta-Analysis Publication Bias – Funnel Plots



Heterogeneity	
Q	1.42
p_Q	0.841
I^2	0.00%
T^2	0.00
T	0.00



Heterogeneity	
Q	4.19
p_Q	0.380
I^2	4.62%
T^2	0.00
T	0.05