Case Study 1: An Evidence-Based Practice Review Report

Theme: School/Setting Based Interventions for Children and young people with special educational needs (SEN).

How effective is the Incredible Years Therapeutic Dinosaur School Programme at promoting positive social, emotional and behavioural outcomes in children with special educational needs?

Summary

This literature review aims to evaluate the effectiveness of the Incredible Years Therapeutic Dinosaur School Programme (Dinosaur School) as an intervention to promote positive social, emotional and behavioural outcomes for children with special educational needs (SEN). The Therapeutic Dinosaur School programme is a small group intervention which aims to address behavioural and emotional problems and promote emotional and social competence in young children (Webster-Stratton, 2016).

A systematic literature review was carried out on 5 databases (PsycINFO, ERIC, Medline, SCOPUS and Web of Science) along with an ancestral search, a search of the Early Intervention Foundation's Guidebook on the Incredible Years Dinosaur School Programme and a search of the Incredible Years website. Six studies were found to meet the inclusion criteria and were reviewed using an adapted version of the Kratochwill coding protocol (UCL Educational Psychology Literature Review Coding Protocol, adapted from the APA Task Force Coding Protocol, 2003) or Horner's Quality Indicators (2005), and were evaluated using Gough's (2007) Weight of Evidence Framework.

The studies reviewed indicated that the Dinosaur School intervention has an effect on behavioural and emotional outcomes for children with ADHD, ODD or CD. However, on the whole, findings were inconsistent and further research is recommended to strengthen and consolidate the evidence base for this intervention.

Introduction

What is it?

The Incredible Years (IY) Therapeutic Dinosaur School Programme, also known as Child Training, Child Therapy or Dina Dinosaur Social Emotional Skills and Problem-Solving Curriculum, is one of a series of IY programmes developed by Webster-Stratton to promote positive social, emotional and behavioural outcomes in a number of ways; including increasing attention, patience, compliance, persistence and emotional regulation and reducing aggression and challenging behaviours (Webster-Stratton, Reid & Beauchaine, 2011). This series of "interlocking programmes" (Hutchings, Bywater, Daley & Lane 2007) includes parent, teacher and child programmes which can be implemented separately or in combination.

The IY Dinosaur School was developed to teach children aged between 4 and 8 problem solving skills and appropriate social behaviours (Hutchings et al., 2007), and is subdivided into two forms of the programme, a universal classroom approach and a small-group therapeutic intervention programme. The therapeutic programme can take place in either a clinic setting or within a school, to be delivered by two group leaders. Group leaders may be

psychologists, therapists, counsellors, social workers or teachers who have undertaken training in the programme (Webster-Stratton, 2016). The programme consists of 18-22 weeks of weekly 2-hour sessions and focuses on several dinosaur characters who attend 'dinosaur school' where Dina Dinosaur is the head teacher. Programme delivery includes puppet interactions, videos, coached play and circle time activities and children are rewarded with 'dinosaur chips' for positive behaviours related to their treatment outcomes. In addition to the sessions themselves children undertake weekly homework tasks with their parents, while teachers and parents are kept updated regularly on the sessions and invited to participate in the planning and application of strategies to support the sessions (EIF, 2017). The curriculum covers 7 units and each addresses a number of key objectives (Table 1) using a range of resources including DVDs, manuals, stickers and visual resources such as feelings cards and posters.

Table 1: Units and objectives of the Incredible Years Dinosaur School Programme

Unit	Unit Objectives
Apatosaurus Unit *"Making friends and learning school rules"*	- Understanding the importance of rules - Participating in the process of rule making - Understanding what will happen if rules are broken - Learning how to earn rewards for good behaviours - Learning to build friendships
Triceratops Unit *"Understanding and detecting feelings"*	- Learning words for different feelings - Learning how to tell how someone is feeling from verbal and nonverbal expressions - Increasing awareness of nonverbal facial communication used to portray feelings - Learning different ways to relax - Understanding why different feelings occur - Understanding feelings from different perspectives - Practising talking about feelings

Unit	Unit Objectives
Iguanodon Unit *"How to be successful in school"*	- Learning how to listen, wait, avoid interruptions and put up a hand to ask questions in the classroom - Learning how to handle other children who poke fun and interfere with the child's ability to work at school - Learning how to stop, think and check work first - Learning the importance of cooperation with the teacher and other children - Practising concentrating and classroom skills
Stegosaurus unit *"Wally teaches Problem Solving Steps"*	- Learning how to identify a problem - Thinking of solutions to hypothetical problems - Learning verbal assertive skills - Learning how to inhibit impulsive reactions - Understanding what apology means - Thinking of alternative solutions to problem situations such as being teased and hit - Learning to understand that solutions have different consequences - Learning how to critically evaluate solutions
Tyrannosaurus Rex unit *"Tiny Turtle teaches anger management"*	- Recognising that anger can interfere with problem solving - Understanding Tiny Turtle's story about managing anger and getting help - Understanding when apologies are helpful - Recognising anger is okay to feel 'inside' but not to act out by hitting or hurting someone else - Learning to control angry reactions - Understanding that things that happen to them are not necessarily hostile or deliberate attempts to hurt them - Practising alternative responses to being teased, bullied or yelled at by an angry adult - Learning skills to cope with another person's anger
Allosaurus Unit *"Molly Manners teaches how to be friendly"*	- Learning what friendship means and how to be friendly - Understanding ways to help others - Learning the concept of sharing and the relationship between sharing and helping - Learning what teamwork means - Understanding the benefits of sharing, helping and teamwork.

Unit	Unit Objectives
	- Practising friendship skills
Brachiosaurus Unit *"Molly explains how to talk with friends"*	- Learning how to ask questions and tell something to a friend - Learning how to listen carefully to what a friend is saying - Understanding why it is important to speak u about something that is bothering you - Understanding how and when to give an apology or compliment - Learning how to enter into a group of children who are already playing - Learning how to make a suggestion rather than give commands - Practising friendship skills

Taken from http://www.incredibleyears.com/programs/child/dinosaur-curriculum/

Psychological Basis

The developer of the Incredible Years series of programmes, describes the multi-faceted theoretical background of the approaches as being fundamentally a developmental approach (Webster Stratton, 2016) drawing on Piaget's theory of cognitive development in addition to other approaches, with a focus on the importance of adult-child relationships.

This includes theories of social learning such as Patterson's work on childhood aggression and deviance (Patterson, DeBaryshe, & Ramsey,1989). Patterson conceptualises a developmental framework for antisocial behaviours, whereby early experiences of poor discipline and monitoring cause child conduct problems which in turn leads to negative outcomes such as academic failure, peer rejection, ongoing deviance and later delinquency (Patterson et al., 1989). According to Drugli and Larsson (2006) child training in social skills is

a valuable component to treating conduct problems, which may relate to the importance of peer rejection in Patterson's model. Bandura's social learning theory is also influential, in particular the role of modelling and motivation in learning and the importance of agency and self-efficacy in behavioural change (Bandura, 1986 & 1999).

Relationship theories including Bowlby (1980) and Ainsworth's (1974) theories of attachment are also influential, which both highlight the importance of experiencing secure attachments in early childhood development and the link between attachment and behaviour. Bowlby describes an internal working model through which our early attachments influence our future relationships, behaviours and understanding of the world (Bowlby, 1980). Geddes (2006) argues that a child's attachment model or style has a significant impact on their performance, behaviour and emotional well-being at school and can present a challenge to teachers.

In seeking to address negative thoughts, feelings and behaviours in the target children Webster-Stratton (2016) also describes taking a cognitive-behavioural approach to intervention, citing cognitive therapy approaches for emotional (Beck, 1979) and behavioural problems (Beck, 2005) as key influences. The IY approach supposes that developing emotion-regulation, emotional literacy and problem solving skills will reduce peer conflict and aggression and increase prosocial behaviours and pro-academic skills (EIF, 2017).

Rationale and relevance to Educational Psychology practice

The importance of addressing the social, emotional and mental health (SEMH) needs of children and young people has emerged as a key priority area for schools, Local Authorities and Educational Psychology Services over recent years. Notably the SEND Code of Practice (2015) introduced a shift from terminology from behavioural, emotional and social difficulties (BESD) to SEMH highlighting a move away from focusing on overt 'challenging' behaviour, to looking behind the behaviour at the needs or difficulties of the individual. Similarly a recent government green paper on mental health in schools states that schools should "offer individual and group help to young people with mild to moderate mental health issues including anxiety, low mood and behavioural difficulties" (DfE & DoH, 2017).

According to Egger and Angold (2006) the five most common childhood psychiatric disorders are ADHD, oppositional defiant disorder, conduct disorder, anxiety disorders and depressive disorders. Children who experience such conditions often exhibit behavioural problems, or have deficits in their social and emotional skills and "struggle to engage in the educational curriculum" (Hutchings et al., 2011), are at increased risk of negative future outcomes such as poor educational attainment (DfE, 2016), school drop-out or exclusion, substance misuse (Egger & Angold, 2006) and an increased risk of criminal activity (DfE & DoH, 2017). Similarly the DfE states that social skills deficits in early childhood can influence further emotional and behavioural disorders in childhood and adolescence (DfE, 2016).

Webster-Stratton argues that intervention in the early years or young childhood can prevent these potential negative outcomes, reduce the development of behavioural problems and improve social and emotional competence (Webster-Stratton, 2016). Early intervention is also supported by the SEND Code of Practice (2015). Given the pertinence of addressing SEMH needs within schools and settings it is important to ensure that current practice, and recommended interventions are evidence-based and effective. This is particularly relevant to Educational Psychologists who are called upon to support schools and settings in the recommendation, training and delivery of evidence-based interventions to meet the needs of children and young people with SEN (Frederickson, 2002). Gough, Oliver and Thomas (2013) argue that systematic reviews of research are important in establishing what is already known and how and what more needs to be learnt in order to inform further research, policy and practice.

The IY Dinosaur School programme is identified by the Early Intervention Foundation (2017) as an evidence based programme. In 2009 Educational Psychologists in the Welsh County of Gwynedd coordinated the introduction of the IY teacher and classroom-dinosaur programmes in all of the county's 102 primary schools and subsequently began to pilot the IY therapeutic dinosaur school programme after recognising that some children needed additional support (Hutchings et al., 2011).

However a 2014 review of the independent research of the IY series highlighted that the IY series of interventions are financially expensive and the main body of research to date has been undertaken either by researchers affiliated with the Incredible Years programme or has focused on combining the Dinosaur School programmes with other elements of the IY series (Pidano & Allen, 2014). Due to the cost, training and time implications of undertaking the IY programmes this current review aims to explore whether the evidence supports the use of the dinosaur school as a therapeutic intervention for schools whether, or not, they also implement other elements of the IY series.

Review Question:

This literature review will address the question: How effective is the Incredible Years Dinosaur School Programme at promoting positive social, emotional and behavioural outcomes in children with special educational needs?

Critical Review of the Evidence Base

Literature Search

A comprehensive literature search was conducted using five electronic databases in January 2018. The databases and search terms can be seen in Table 1 below. An additional search was also carried out on the Incredible Years website and an Early Intervention Foundation Guidebook on the intervention (EIF, 2017), in addition to an ancestral search.

Table 2: Database Search Terms

Database	Search Terms
PsycINFO ERIC Medline Web of Science Scopus	(1) “Dinosaur School” OR “Dina Dinosaur” (2) “Dinosaur Curriculum” (3) “Incredible Years” AND classroom (1) OR (2) OR (3)

A flowchart depicting the literature search and screening process can be seen in Figure 1, while the inclusion and exclusion criteria used for screening the articles is listed in Table 2. Following the screening process six studies were found to meet the criteria for inclusion in the review, these are listed in Table 3 while the studies excluded following full-text screening are listed in Appendix 1 along with the criteria by which they were excluded.

Table 3: Inclusion and Exclusion Criteria

Criteria	Inclusion	Exclusion	Rationale
1. Type of study and data	Reports original or previously unreported quantitative findings of an empirical study with experimental or quasi-experimental design.	Non experimental or empirical study. Findings are solely qualitative or the study reports the same data as another study included in the review.	Replication of the same findings could cause a bias in the review’s findings. Quantitative data allows for more direct comparison of effectiveness between studies.

Criteria	Inclusion	Exclusion	Rationale
2. Language	Study must be published in English	Study published in a language other than English	Reviewer does not have the means to translate articles for review
3. Participants	Participants must be of primary or pre-school age and have identified SEN or identified factors putting them at higher risk of SEN.	Participants not of primary or pre-school age. Participants don't have SEN or identified risk factors.	The intervention is designed for this age group and the reviewer seeks to review the efficacy of the intervention undertaken with children with SEN.
4. Intervention	Experimental group must receive the complete Dinosaur School Therapeutic Intervention or not more than 1 week less of the complete intervention. The Intervention must not be combined with more than 1 other element of the Incredible Years series.	Other intervention, universal dinosaur school, more than 1 other element of the IY series or more than 1 week less than the recommended duration of dinosaur school intervention.	Fidelity to the intervention is important in reviewing efficacy and enabling a fair comparison between studies.
5. Measures	Pre and post intervention measures assessing social/behavioural and emotional factors	Study does not have pre and post intervention measures or does not measure social/behavioural and emotional outcomes.	To establish a direct measure of the effectiveness of the intervention in relation to the target variables.

Figure 1: A Flow Diagram of the Literature Search Process

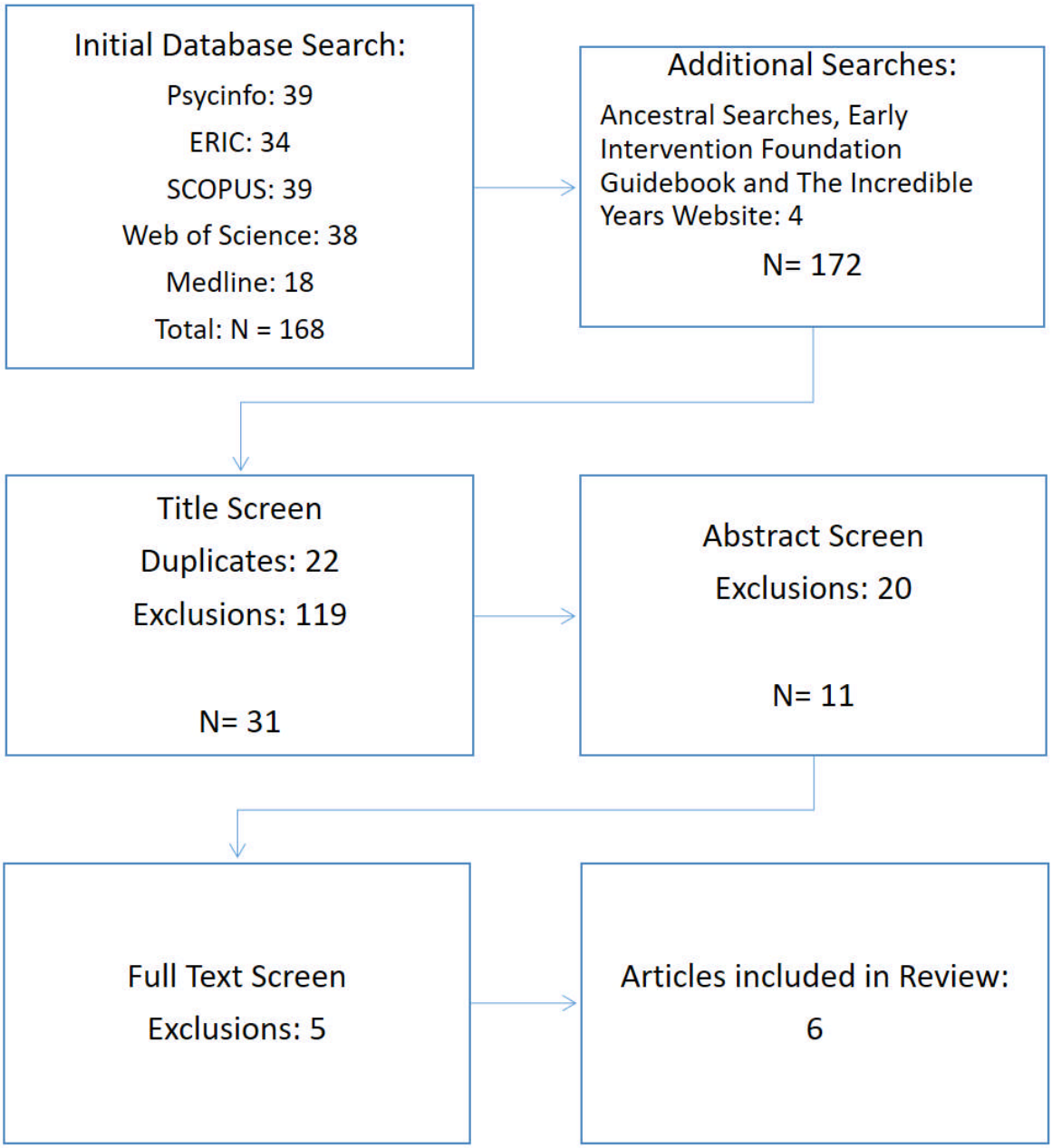

Table 4: Studies Included in the Review

	Study Reference
1.	Larsson, B., Fossum, S., Clifford, G., Drugli, M. B., Handegård, B. H., & Mørch, W.-T. (2009). Treatment of oppositional defiant and conduct problems in young Norwegian children. *European Child & Adolescent Psychiatry*, *18*(1), 42–52.
2.	Webster-Stratton, C. H., Reid, M. J., Beauchaine, T., (2011). Combining Parent and Child Training for Young Children with ADHD, *Journal of Clinical Child & Adolescent Psychology, 40(2), 191-203.*
3.	Webster-Stratton, C., & Hammond, M. (1997). Treating children with early-onset conduct problems: A comparison of child and parent training interventions. *Journal of Consulting and Clinical Psychology*, *65*(1), 93-109.
4.	Herman, K. C., Borden, L. A., Reinke, W. M., & Webster-Stratton, C. (2011). The Impact of the Incredible Years Parent, Child, and Teacher Training Programs on Children's Co-Occurring Internalizing Symptoms. *School Psychology Quarterly : The Official Journal of the Division of School Psychology, American Psychological Association*, *26*(3), 189–201.
5.	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 & Adolescent Psychology, 33(1), 105-124.*
6.	Hutchings, J., Bywater, T., Daley, D., & Lane, E. (2007). A Pilot Study of the Webster-Stratton Incredible Years Therapeutic Dinosaur School Programme. *Clinical Psychology Forum*, *170*, 21–24.

Appraising the Studies

The six studies selected for review were summarised into a table in order to map the field of research (see Appendix 2) and were then coded using an adapted version of Kratochwill's APA Task Force Coding Protocol (Kratochwill, 2003) for the group based designs and using Horner's Quality Indicators (Horner, 2005) for one study which used a single case design. This information was then used to inform the application of Gough's Weight of Evidence (WoE) Framework (Gough, 2007). This framework was used to enable a more

objective judgement to be given as to the quality and relevance of the studies chosen for review. The WoE framework consists of four components, each of which is given a score of 1 (low), 2 (medium) or 3 (high). Table 5 details the four components of the WoE framework and Table 6 outlines the WoE scores for each study. An example of the completed coding protocols can be seen in Appendix 4 alongside the full WoE criteria in Appendix 3.

Table 5: Weight of evidence A, B, C & D according to Gough's Weight of Evidence Framework (2007)

Weight of Evidence A	**Weight of Evidence B**	**Weight of Evidence C**	**Weight of Evidence D**
Methodological Quality (Integrity of the study)	Methodological Relevance (Appropriateness of research design)	Topic Relevance (Relevance to current review question)	Overall weight of evidence in relation to the current review question. (Average of WoEs A, B and C)

Table 6: Weight of Evidence Scores

Study	WoE A	WoE B	WoE C	WoE D (Overall weight of evidence)
1. *Larsson et al., 2009*	*2.3*	*3*	*1*	*2.1 (medium)*
2. *Webster-Stratton, Reid & Beauchaine, 2011*	*2*	*2*	*1*	*1.7 (medium)*
3. *Webster-Stratton & Hammond, 1997*	*2*	*3*	*2*	*2.3 (medium)*
4. *Herman et al. 2011*	*2*	*2*	*1*	*1.7 (medium)*
5. *Webster-Stratton, Reid & Hammond, 2004*	*2.7*	*3*	*2*	*2.6 (high)*
6. *Hutchings et al. 2007*	*1.3*	*1*	*2*	*1.4 (low)*

Participants

The studies chosen for review were conducted in the United States of America (Webster-Stratton et al., 2011; Webster-Stratton & Hammond, 1997; Webster-Stratton, Reid & Hammond 2004 & Herman et al., 2011), Norway (Larsson et al., 2009) and the United Kingdom (Hutchings et al., 2007). In total 500 participants were included in this review aged between 4 and 11 years, 81.6% of the participants were male. Participants were chosen for all studies due to clinical diagnoses associated with social, emotional and behavioural difficulties, the diagnoses included in the review were Oppositional Defiant Disorder (ODD), Conduct Disorder (CD) and Attentional Deficit Hyperactive Disorder (ADHD). Some participants had comorbid diagnoses of ODD and CD (Larsson et al., 2009 & Webster-Stratton & Hammond, 1997) or CD and ADHD (Hutchings et al., 2007).

Sample sizes between the studies were varied, as were the sizes of individual group conditions and the target intervention groups ranged from 9 to 52 participants with comparison group sizes ranging from 22 to 51 participants. Cohen (1992) identifies minimum group sizes depending on the effect size, and recommends that where there are at least two groups these sizes refer to each group, suggesting that the group based designs were all underpowered for at least some measures (see tables 7 and 8). This was reflected in the scoring of WoE A as group based designs had to have an adequate sample size in all groups in order to achieve a 'high' weighting for analysis within the methodological quality criteria (see Appendix 3).

Design

Of the six articles included in the review two reported on the same study, using different findings (Herman et al., 2011 & Webster-Stratton, et al., 2004) therefore for the purposes of the review their design will be described as one study. All studies used a Randomised Control Trial design apart from one (Hutchings et al., 2007) which used a single case design, this means that all but one of the studies included at least one comparison group. The group design studies all included a waitlist control condition for comparison, as this is argued to be more ethical than a no-intervention condition as no participants are denied access to an intervention which could be effective for them (Barker, Pistrang & Elliot, 2015). Other than Webster-Stratton et al. (2011) all other group-design studies also included at-least one active comparison group, wherein participants were exposed to an alternative IY programme, including

parent or teacher training programmes. In comparing these studies the use of an active comparison groups was valued higher in WoE A.

The single case design study (Hutchings et al., 2007) was a pilot study of the intervention programme. Horner (2005) describes the importance of single case design studies meeting certain experimental criteria, including using baseline measures to enable comparison and within-subject control. Horner suggests a multiple baseline design of at least 5 points of measurement to establish a pattern of response, similarly it is recommended that experimental effect is demonstrated with measurements at three different points in time (Horner, 2005). However the pilot study included in the review does not report multiple time points of measurement at pre or post intervention, although there are multiple measures reported. This is reflected in the judgement of WoE A for this study.

All studies took pre and post intervention measures and three also reported follow up findings (Larsson et al., 2009; Webster-Stratton & Hammond, 1997 & Herman et al., 2011 & Webster-Stratton, et al., 2004). Across all studies the majority of measures were quantitative, though four articles also reported some qualitative findings including consumer (parental or teacher) satisfaction with intervention (Larsson et al., 2009; Webster-Stratton et al., 2011; Webster-Stratton & Hammond, 1997 & Webster-Stratton et al., 2004). For the purposes of this study only quantitative measures relating to child outcomes are included.

Quality of Intervention

The quality of the interventions have been considered according to their fidelity to the intervention, the relevance of implementation to the current review and the measures used, which were also reflected in the WoE C scores.

All interventions were facilitated by group leaders who had undertaken IY training although none of the studies in the current review were undertaken by trained teachers, nor were they undertaken within the participants' school setting although the intervention is designed to be school-based. The intervention setting was included in WoE C and to achieve a 'high' rating for topic relevance the study would have had to take place within a school. Most studies included teacher reported outcomes, which are relevant to considering their efficacy for schools. However Larsson et al.'s study only gives a summary of generalization effects to school and does not directly report the findings. Additionally Herman et al.'s (2011) article does not report any teacher rated findings.

Although all studies reported steps taken to ensure fidelity to the intervention programme it is worth noting that the designs varied in the duration of intervention undertaken. Hutchings et al. (2007) implemented only 17 weeks of the intervention, 1 week less than recommended for the programme, while all others were within the recommended 18 - 22 weeks. This was deemed to reduce the level of fidelity to the intervention and as a result Hutchings et al. scored lower for WoE C. This variation in intervention duration may be significant in considering the overall effect of the interventions. For the studies that reported participant attendance there was also variation in the amount of

sessions attended (see appendix 2), with some participants identified as attending as few as 13 sessions (Larrson et al., 2009; Webster-Stratton & Hammond, 1997).

Similarly three of the studies combined the Dinosaur School with another IY programme in the intervention condition (Larsson et al., 2009; Webster Stratton & Hammond, 1997; Webster-Stratton et al., 2011). This makes it more difficult to establish whether the outcomes in these studies are due to the application of the Dinosaur School programme and as a result has been reflected in WoE scores for topic relevance. However this is helpful in enabling comparative data which is beneficial in considering the issue of whether the programme is effective, and likely to be worth investing in, as a standalone intervention for schools or whether it's efficacy is cumulative as part of the interlocking IY programmes.

All studies outlined the measures used and followed the IY Dinosaur School manuals, meaning they should be replicable. Some of the behavioural measures were fairly consistent across studies, for example all used the Eyberg Child Behaviour Inventory (Robinson, Eyberg & Ross, 1980) – an inventory of conduct problems which provides two summary scores, Problems (the number of reported conduct problems) and Intensity (the intensity of these problems) (Webster-Stratton et al., 2011). Four of the studies also used the Child Behaviour Checklist (Achenbach, 1991) which can measure both externalising and internalising symptoms. Other measures used were more varied and included the Conner's teacher and parent rating scales (Conners, 1998 & Conners, Sitarenious, Parker, & Epstein, 1998), the Preschool Behaviour

Questionnaire (Behar, 1977), the Self Control Rating Scale (Kendall & Wilcox, 1979) and observation data. Measures of emotional outcomes included the Strengths and Difficulties Questionnaire (Goodman, 1997) and the Wally Problem Solving and Feeling Tests (Webster-Stratton, 1990), unpublished assessments developed by the Incredible Years team. The relevance of measures used was incorporated into WoE C and their validity was incorporated into WoE A.

The majority of measures used were based on parent report, which could be considered to be a limitation in their relevance to considering school or setting based efficacy, and also may be influenced by some parents participating in parent training as parents may rate progress higher if they also received intervention. This can, to some extent, be controlled for by considering findings reporting on combined interventions against the child intervention alone. Where only one source of assessment was measured, for example only parent sourced measures were undertaken, this was reflected in WoE A.

Findings

In the group-design studies effect sizes were either given or were calculated using Cohen's d (Cohen, 1992). Cohen's d was determined by calculating the difference between the reported means of the two groups (intervention and control), and dividing by the pooled standard deviation. A small group pilot study by Hutchings et al. (2007) identified that Cohen's effect sizes had been calculated; "Using Cohen's 1988 guidelines for effect sizes, a clinically significant improvement was achieved for all… measures" (Hutchings et al., 2007). However effect sizes were not reported, and limited information was

given to recalculate, meaning that the effect sizes for this study were calculated using Lipsey and Wilson's (2001) protocol. Table 6 outlines effect sizes for key outcomes.

On the whole the findings suggested that the intervention had an effect on behavioural and emotional outcomes (see Table 7). However these findings varied between, and even within, studies with effect sizes for different measures ranging from no effect to large effect (e.g. Webster-Stratton & Hammond, 1997), with no effect found by Webster-Stratton and Hammond (1997) on their teacher reported outcome measure; the Preschool Behaviour Questionnaire. In comparing the findings of these studies it is important to be mindful of their WoE ratings in order to ensure the evidence is considered with a view to its quality and relevance to the current review. For example Hutchings et al. (2007) showed the most consistent effect sizes across measures with one moderate and all others large, however this study received the lowest WoE D score (for overall weight of evidence). The study with the highest WoE D rating reported a significant improvement in almost all measures compared to control however effect sizes for school and peer related measures were small.

Of the group design studies those with the largest effect sizes (Webster-Stratton et al., 2011 & Webster-Stratton & Hammond, 1997) had the longest intervention periods at 20 and 22 weeks respectively while the intervention in the former was combined parent and child. It is important to consider whether additional time, or additional intervention could be a key factor in these findings. It is also important to consider the outcomes in relation to the specific difficulties of the participants, for example Webster-Stratton and Hammond

(1997) found moderate effects on the CBCL and large effects on the ECBI-intensity measures with both mother and father reports for participants with ODD/CD. However Larsson et al.'s (2009) participants also had diagnoses of ODD/CD and Mother reported behaviour problems (ECBI) and aggression (CBCL) found no effect comparing the parent and child combined condition with the parent alone, but found a moderate effect comparing the intervention condition with control. This raises the question as to whether there was value added by the Dinosaur School in this instance. While a combined intervention study focusing on participants with ADHD (Webster-Stratton et al., 2011) found no effect in teacher reported inattentiveness and a small effect in hyperactivity (both CTRS-R).

In considering emotional outcomes the evidence was also inconsistent, with CBCL measures of internalising symptoms finding Moderate (Larsson et al., 2009) (mother report) to small (Webster-Stratton et al., 2011) (mother and teacher) or no effect (father report). The latter study did report a large effect for both parental SCS measures in emotion regulation and a large effect was found for parent and teacher SDQ measures by Hutchings et al. (2007).

Table 7: Effect Sizes for Primary Outcomes

Study	Sample Size	Measure	Effect Size	Effect Size Descriptor	Overall WoE
Larsson et al., 2009	55 in Target intervention group (PT+CT)	Mother Report:			**Medium**
		ECBI – Intensity	PT vs PT + CT d = 0.22	Small	
			PT + CT vs WLC d = 0.42	Small	
	51 in alternative intervention group (PT)	EBCI – Problems	PT vs PT + CT d = - 0.03	No effect	
			PT + CT vs WLC d = 0.55	Moderate	
	Control Group: 30	CBCL – aggression	PT vs PT + CT d = -0.15	No effect	
			PT + CT vs WLC d = 0.75	Moderate (high)	
		CBCL – attention	PT vs PT + CT d = - 0.03	No effect	
			PT + CT vs WLC d = 0.59	Moderate	
		CBCL - internalising	PT vs PT+CT d = -0.05	No effect	
			PT + CT vs WLC d = 0.6	Moderate	

Study	Sample Size	Measure	Effect Size	Effect Size Descriptor	Overall WoE
		Father Report:			
		ECBI – Intensity	PT vs PT + CT d = 0.54	Moderate	
			PT + CT vs WLC d = 0.17	No effect	
		EBCI – Problems	PT vs PT + CT d = 0.67	Moderate	
			PT + CT vs WLC d = 0.02	No effect	
		CBCL – aggression	PT vs PT + CT d = - 0.12	No effect	
			PT + CT vs WLC d = 0.52	Moderate	
		CBCL – attention	PT vs PT + CT d = - 0.24	Small	
			PT + CT vs WLC d = 0.50	Moderate	
		CBCL - internalising	PT vs PT+CT d = -0.36	Small	
			PT + CT vs WLC d = 0.33	Small	
Webster-Stratton, Reid & Beauchaine, 2011	Intervention Group (PT+CT): 49	Mother Report:			**Medium**
		ECBI-Intensity	ηp^2 .22	Large	

Study	Sample Size	Measure	Effect Size	Effect Size Descriptor	Overall WoE
		CBCL – Aggression	ηp^2 .04	Small	
	Control Group: 50	CBCL – Attention	ηp^2 .04	Small	
		CBCL - Externalising Broadband	ηp^2 .06	Moderate	
		CBCL - Internalising	ηp^2 .02	Small	
		CPRS-R Oppositional	ηp^2 .11	Moderate	
		CPRS-R Inattentive	ηp^2 .07	Moderate	
		CPRS-R Hyperactive	ηp^2 .13	Moderate	
		SCS – Emotion Regulation	ηp^2 .22	Large	
		SCS – Social Competence	ηp^2 .17	Large	
		Father Report:			
		ECBI-Intensity	ηp^2 .16	Large	
		CBCL – Aggression	ηp^2 .05	Small	
		CBCL – Attention	ηp^2 .03	Small	

Study	Sample Size	Measure	Effect Size	Effect Size Descriptor	Overall WoE
		CBCL – Externalising Broadband	ηp^2 .06	Moderate	
		CBCL - Internalising	ηp^2 <.01	No effect	
		CPRS-R Oppositional	ηp^2 .05	Small	
		CPRS-R Inattentive	ηp^2 .06	Moderate	
		CPRS-R Hyperactive	ηp^2 .06	Moderate	
		SCS – Emotion Regulation	ηp^2 .24	Large	
		SCS – Social Competence	ηp^2 .12	Moderate	
		Teacher Report:			
		TRF – Externalising Broadband	ηp^2 .04	Small	
		TRF- Internalising Broadband	ηp^2 .03	Small	
		CTRS-R Oppositional	ηp^2 .01	Small	
		CTRS-R Inattentive	ηp^2 <.01	No effect	
		CTRS-R Hyperactive	ηp^2 .01	Small	

Study	Sample Size	Measure	Effect Size	Effect Size Descriptor	Overall WoE
		DPICS Behaviour Observations:			
		Child Deviance	ηp^2 .06	Moderate	
		Child Positives	ηp^2 .01	Small	
		Wally Feelings Test	ηp^2 .09	Moderate	
		Wally problem Solving Test	ηp^2 .04	Small	
Webster-Stratton & Hammond, 1997	Intervention Group: n = 27	Mother Report			**Medium**
	27 mothers, 20 fathers	CBCLB	d=-0.50	Moderate	
		ECBI-I	d= -1.34	Large	
	WLC: n = 22	Father Report			
	22 mothers, 18 fathers	CBCLB	d=-0.71	Moderate	

Study	Sample Size	Measure	Effect Size	Effect Size Descriptor	Overall WoE
		ECBI-I	d=-0.84	Large	
		Teacher Report			
		PBQ	d=0.18	No Effect	
Herman et al. 2011		Mother Report			**Medium**
		CBCL	d=-0.09	No Effect	
Webster Stratton, Reid & Hammond, 2004		Mother Report			**High**
		CCP – at Home	d= -0.64	Moderate	
		Father Report			
		CCP – at Home	d=-0.61	Moderate	
		CSC with peers	d=0.35	Small	
		CCP at School	d=-0.30	Small	

Study	Sample Size	Measure	Effect Size	Effect Size Descriptor	Overall WoE
Hutchings et al. 2007	N = 9	Parent Report			**Low**
		ECBI-I	d=-1.05	Large	
		ECBI-p	d=-0.72	Moderate	
		SDQ	d=-0.93	Large	
		SDQ Impact	d=-1.98	Large	
		SCRS	d=-1.47	Large	
		Teacher Report			
		SDQ	d=-2.05	Large	
		SCRA.	d=-2.02	Large	
			Effect sizes were not reported and limited data was given, effect sizes were calculated using Lipsey & Wilson (2001)		

Table 8: Effect size descriptors and recommended group size

Effect Size	Small	Medium	Large
Cohen's d (d) (Cohen, 1992)	0.2	0.5	0.8
Partial eta squared (ηp^2) (Cohen, 1988)	0.01	0.06	0.14
Recommended group size (Cohen, 1992)	393	64	26

Significance set at 0.05, power set at 0.8

Conclusion and Recommendations

This review aimed to review the efficacy of the Dinosaur School intervention in children with SEN, particularly within an education setting. Six studies were included, appraised and discussed. One received a 'low' WoE D rating (Hutchings et al., 2007) and one 'high' (Webster-Stratton et al., 2004) while the others received 'medium'. Although the studies in this review showed some promising findings to suggest the Dinosaur School is effective in promoting positive SEMH outcomes for children with SEN, the findings were inconsistent both across and within studies and are also subject to limitations.

These limitations have been described and incorporated into Weight of Evidence appraisal, as outlined in table 6. This includes the lack of consistency in intervention duration and also in the outcome measures used and

particularly the sources used for these measures. The majority were parent reported measures while this review was interested in considering the relevance and efficacy of the intervention for schools. In addition to the outcomes not focusing sufficiently on school the studies were not implemented by teachers, only one took place within the UK and they took place in clinical rather than school settings, which enables a greater level of experimental control. However these factors limit the relevance and usefulness of the current research base in considering efficacy for school based intervention which suggests more may need to be done in order for Educational Psychologists to recommend the programme for practice in schools at this time

There was also a lack of comparison to other intervention programmes, beyond the IY series, which would be helpful in evaluating how effective separate programmes are at promoting the target outcomes and may give useful information for schools and EP services. Similarly the focus of research at present seems to focus on a small area of SEN with the evidence base searched and reviewed seeming to focus on just three diagnoses. Based on this limitation to the evidence base the programme should only be recommended for the children and young people with these same diagnoses and with other limitations to the evidence also in mind. This may limit the likelihood of it being recommended in practice or, if recommended, being taken up by schools as the target population may represent a very small subset of the student population in a given school.

In reviewing the included studies it is also important to highlight that the majority of studies were undertaken by researchers affiliated with the IY programme. In a previous review of the IY series the “paucity” of independent

research is highlighted (Pidano & Allen, 2014). Although this past review did not focus on the dinosaur school their conclusions are similar in also reflecting on the relatively high cost of training and implementation for the IY programmes in relation to their evidence base. Drugli and Larsson (2006) argue that treatments targeting the home and school are more likely to produce effective behavioural change compared to parent intervention alone. However the logistical and financial demand of intervention is also greater if the dinosaur programme is combined with other IY programmes such as the parent intervention and there would need to be a greater research focus on the outcomes seen in schools.

With this in mind the question remains as to whether the Dinosaur School intervention is cost effective and practically feasible for EPs or schools to train in and implement, leading to the recommendation for more research to be undertaken to address the current limitations to the evidence base. In particular there is a need for school based research within the UK and for research to be carried out on a more diverse range of students with SEN. Future studies would benefit from ensuring they have large enough sample sizes within each group to be confident in the generalisability of findings and to ensure sufficient power. It would also be beneficial for active comparison against an alternative SEMH intervention in order to make more robust conclusions about whether the IY Therapeutic Dinosaur School intervention can be considered an effective, evidence based approach for SEN children exhibiting social, emotional or behavioural difficulties.

References and Appendices

References:

Achenbach, T. (1991). *Manual for the teacher's report form and teacher version of the child behaviour profile.* Burlington: University of Vermont, Department of Psychiatry

Ainsworth, M. (1974). Infant–mother attachment and social development: Socialization as a product of reciprocal responsiveness to signals. In M. Richards (Ed.), *The integration of the child into the social world*. Cambridge: Cambridge University Press.

Bandura, A. (1986). *Social foundations of thought and action*. Englewood Cliffs, NJ: Prentice-Hall.

Bandura, A. (1999). Social cognitive theory : An agentic perspective. *Asian Journal of Social Psychology*, *2*, 21–41.

Barker, C., Pistrang, N., & Elliott, R. (2015). *Research methods in Clinical Psychology: An introduction for students and practitioners*. Third Edition. Chichester: Wiley-Blackwell.

Beck, A. T. (1979). *Cognitive therapy and emotional disorders*. New York: New American Library.

Beck, J. S. (2005). *Cognitive therapy for challenging problems*. New York: The Guilford Press.

Behar, L.B. (1977). The Preschool Behaviour Questionnaire. *Journal of Abnormal Child Psychology,* 5, 265-275.

Bowlby, J. (1980). *Attachment and loss: Loss, sadness, and depression*. New York: Basic Books.

Cohen, J. (1988). Statistical power analysis for the behavioural sciences. USA: Lawrence Erlbaum Associates

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155.

Conners, C.K (1998). Revision and restandardisation of the Conners' Teacher Rating Scale (CTRS-R): Factor structure, reliability and criterion validity. *Journal of Abnormal Child Psychology,* 26, 279-291.

Conners, C.K., Sitarenious, G. Parker, J.D.D. & Epstein, J.N. (1998). The revised Conners' Parent Rating Scale (CPRS-R): Factor structure, reliability and criterion validity. *Journal of Abnormal Child Psychology,* 26, 257-268.

Content and objectives of the Incredible Years child training programs (aka Dina Dinosaur social emotional skills and problem-solving curriculum) for ages

4-8. (n.d.) Retrieved from http://www.incredibleyears.com/programs/child/dinosaur-curriculum/

Department for Education (DfE) (2015). Special educational needs and disability code of practice: 0 to 25 years. London: Department for Education.

Department for Education (DfE) (2016) Mental health and behaviour in schools: Departmental advice for school staff. London: Department for Education.

Department for Education (DfE) and Department of Health (DoH), (2017). Transforming children and young people's mental health provision. London:

Drugli, M. B. & Larsson, B. (2006). Children aged 4-8 years treated with parent training and child therapy because of conduct problems: generalisation effects to day-care and school settings. *European Journal of Adolescent Psychiatry,* 15, 392-399.

Early Intervention Foundation (EIF), (2017). EIF Guidebook: Incredible Years Child Training (Dinosaur School). Downloaded from: http://guidebook.eif.org.uk/programme/incredible-years-child-training-dinosaur-school

Egger, H. L., & Angold, A., (2006). Common emotional and behavioral disorders in preschool children: Presentation, nosology, and epidemiology. *Journal of Child Psychology and Psychiatry*, *47*, 313–337

Frederickson, N., (2002) Evidence Based Practice and Educational Psychology. *Educational and Child Psychology*, 19 (2) pp. 96-111

Geddes, H. (2006), Attachment in the Classroom. Pastoral Care in Education, London: Worth Publishing.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38, 581-586.*

Gough, D., (2007). Weight of Evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, *22*(2), 213–228.

Gough, D., Oliver, S. & Thomas., J. (2013). Learning from Research: Systematic reviews for informing policy decisions. A quick guide. *The Alliance for Useful Evidence,* 1-38.

Harden, A., & Gough, D., (2012). Quality and Relevance Appraisal. In D. Gough, S. Oliver, & J. Thomas (Eds.), *An Introduction to Systematic Reviews* (pp. 153–178).London: Sage.

Herman, K. C., Borden, L. A., Reinke, W. M., & Webster-Stratton, C. (2011). The Impact of the Incredible Years Parent, Child, and Teacher Training

Programs on Children's Co-Occurring Internalizing Symptoms. *School Psychology Quarterly : The Official Journal of the Division of School Psychology, American Psychological Association*, *26*(3), 189–201

Horner, R.H., (2005). The use of single-subject research to identify evidence-based practice in special education. *Council for Exceptional Children,* 71(2), 165-179.

Hutchings, J., Bywater, T., Daley, D., & Lane, E. (2007). A Pilot Study of the Webster-Stratton Incredible Years Therapeutic Dinosaur School Programme. *Clinical Psychology Forum*, *170*, 21–24

Hutchings, J. Bywater, T., Gridley, N., Whitaker, C.J., Martin, P. & Gruffyd, S. (2011). The incredible years therapeutic social and emotional skills programme: A pilot study. *School Psychology International,* 33(3), 285-293.

Kendall, P.C. & Wilcox, L.E. (1979). Self-control in children: Development of a rating scale. *Journal of Consulting and Clinical Psychology,* 47, 1020-1029.

Kratochwill, T. R. (2003). Evidence-Based Practice: Promoting Evidence-Based Interventions in School Psychology. *School Psychology Quarterly, 18*(4), 389 - 408.

Larsson, B., Fossum, S., Clifford, G., Drugli, M. B., Handegård, B. H., & Mørch, W.-T. (2009). Treatment of oppositional defiant and conduct problems in young Norwegian children. *European Child & Adolescent Psychiatry*, *18*(1), 42–52.

Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Patterson, G.R., DeBaryshe, B. & Ramsey, E. (1989). A developmental perspective on antisocial behaviour. *American Psychologist,* 44, 329-225.

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.

Robinson, E.A., Eyberg, S.M. & Ross, A.W. (1980). The standardisation of an inventory of child conduct problem behaviours. *Journal of Clinical Child Psychology,* 9, 22-28.

Slavin, R., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis*, *31*(4), 500-506.

Webster-Stratton, C. (1990). Wally Feelings Test. *Unpublished manuscript*, University of Washington, Seattle.

Webster-Stratton, C. (1990) Wally Game: A problem-solving test. *Unpublished manuscript,* University of Washington, Seattle.

Webster-Stratton, C. (2016). The Incredible Years® Series: A Developmental Approach. In VanRyzin, M., Kumpfer, k., Fosco, G. & Greenburg, M. *Family-Centered Prevention Programs for Children and Adolescents: Theory, Research and Large-Scale Dissemination.* New York, Psychology Press.

Webster-Stratton, C., & Hammond, M. (1997). Treating children with early-onset conduct problems: A comparison of child and parent training interventions. *Journal of Consulting and Clinical Psychology, 65*(1), 93-109.

Webster-Stratton, C. & Reid, M.J. (2008). Adapting the Incredible Years child dinosaur social, emotional and problem-solving intervention to address comorbid diagnoses. *Journal of Children's Services, 3(3), 17-30.*
Webster-Stratton, C. H., Reid, M. J., Beauchaine, T., (2011). Combining Parent and Child Training for Young Children with ADHD, *Journal of Clinical Child & Adolescent Psychology*, *40(2), 191-203.*

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 & Adolescent Psychology, 33(1), 105-124*

Webster-Stratton, C. H., Reid, M. J., Beauchaine, T., (2011). Combining Parent and Child Training for Young Children with ADHD, *Journal of Clinical Child & Adolescent Psychology*, *40(2), 191-203.*

Appendix 1: List of Excluded Studies

Appendix 1

Excluded Studies	Reason for Exclusion
1. Hutchings, J., Lane, E., Owen, R. E., & Gwyn, R. (2004). The introduction of the Webster-Stratton Classroom Dinosaur School Programme in Gwynedd, North Wales : A pilot study. *Educational and Child Psychology, 21(4).*	Criteria 3
2. Baker-Henningham, H., Walker, S., Powell, C., & Gardner, J. M. (2009). 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.	Criteria 3
3. Hutchings, J., Bywater, T., Gridley, N., Whitaker, C. J., Martin-Forbes, P., & Gruffydd, S. (2011). The Incredible Years Therapeutic Social and Emotional Skills Programme: A Pilot Study. *School Psychology International*, *33*(3), 285–293.	Criteria 4
4. Drugli, M.B. & Larsson, B. (2006). Children aged 4-8 years treated with parent training and child therapy because of conduct problems: generalisation effects to day-care and school settings. *European Child & Adolescent Psychiatry*, 15, 392 – 399.	Criteria 1
5. 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(5), 471-488.*	Criteria 4

Appendix 2: Mapping the Field

Appendix 2

#1 Larsson et al. 2009					
Participants	**Intervention**	**Design**	**Comparison Group**	**Measures**	**Reported Outcomes / Key Findings**
136 children identified due to oppositional or conduct problems. Of these 9 withdrew before treatment began and 2 subsequently dropped out from the alternative experimental group. 125 children completed the study. Ages ranged from 4-8 and the mean age of participants	52 children completed the intervention condition. Intervention consisted of Incredible Years Parent training and Dinosaur School. Groups of six children with two therapists for 2 hour weekly sessions over 18 weeks. The Programme was followed with hand-outs and video materials translated into Norwegian.	RCT with three conditions. Parent and Child Therapy (PT + CT), Parent Therapy (PT) and Waitlist Control (WLC).	28 participants were in a waitlist control group were offered treatment after 6 months. An alternative experimental condition was included of 51 participants who received just the parent training (4 withdrew before treatment began and 2 subsequently dropped out).	Eyberg Child Behaviour Inventory (ECBI). Child Behaviour Checklist (CBCL). Parent Practices Inventory (PPI). Parent Stress Index (PSI).	Both the target (PT+CT) and alternative intervention condition (PT) led to a reduction in child behaviour problems compared to the waitlist condition. Children in the target intervention group (PT+CT) showed a greater reduction in measures of aggression, attention difficulties and internalising symptoms than the alternative intervention condition and the waitlist condition. However scores on the Eyberg Child Behaviour Inventory (ECBI) showed

within the experimental group was 6.7 years	4 children attended less than 75% of sessions Parents also received 2-hour weekly training with two therapists over 12-14 weeks.				no effect to small effect sizes in the target condition. PT+CT showed greater generalisation to day care or school settings, but these reports were not sustained at follow up and are not reported in full in the current study. 12 month follow up data was also collected.

#2 Webster-Stratton, Reid & Beauchaine 2011					
Participants	**Intervention**	**Design**	**Comparison**	**Measures**	**Outcomes / Key Findings**
99 children with a diagnosis of ADHD (48 with comorbid Oppositional Defiant Disorder). Ages ranged from 4-6 years.	49 children received the Incredible Years Parent Training and Dinosaur School programs took place at the same time over 20 weeks in two hour weekly sessions of dinosaur school and	RCT with two conditions. Parent and Child Therapy (PT + CT) and Waitlist	50 children were in a waitlist control condition	PPI CBCL Parent CBCL - Teacher Report Form (TRF) ECBI	The intervention condition (PT + CT) showed a reduction in hyperactive, inattentive and oppositional behaviours on a range of measures. For ECBI - Intensity reported scores gave very low standard

	two hours of weekly parent training. Child attendance was not reported.	Control (WLC).		Conners' Parent Rating Scale-Revised (CPRS-R) Conner's Teacher Rating Scale-Revised (CTRS-R) Social Competence Scale Behaviour Observations Dyadic Parent-Child Interactive Coding System-Revised Wally Problem Solving Test Wally Feelings Test Parent Satisfaction Questionnaires	deviations which leads to highly significant findings on this outcome and large effect size.

#3 Webster-Stratton & Hammond, 1997

Participants	Intervention	Design	Comparison Group	Measures	Outcomes / Key Findings

97 children were aged between 4 and 7 years with child misconduct problems who met clinical criteria for Oppositional Defiant Disorder or Conduct Disorder.	27 children were in the target intervention group and received child training (IY Dinosaur School Therapeutic Programme) over 22 weeks. All participants attended 13 or more sessions. All but one attended 15 or more.	RCT with 4 conditions	22 children in waitlist control 26 in an alternative intervention condition received just parent training (PT) 22 in an alternative intervention condition received a combination of parent and child training (PT + CT)	CBCL – total behaviour problems ECBI – Intensity Preschool Behaviour Questionnaire (PBQ) Teacher reported total problems Parenting Stress Index (PSI)	There was an improvement in parent reported behaviour problems, however mothers scoring on the CBCL scale was non-significant in the intervention vs waitlist control condition, while the alternative intervention conditions were significant in this measure. In all measures the target intervention condition performed less well than the parent training intervention. Effect sizes for the intervention condition were found to be moderate to large, with scores on the ECBI-Intensity scale statistically significant at the $p < 0.01$ level for father reported scores and $p < 0.001$ level for mother reported scores, both with a large effect size. PBQ scores showed high variance and were statistically non-significant.

#4 Herman et al. 2011					
Participants	**Intervention**	**Design**	**Comparison Group**	**Measures**	**Outcomes / Key Findings**
159 participants aged 4 to 8 years with a diagnosis of Oppositional Defiant Disorder.	30 children received the target intervention (CT) for 18 – 19 weeks in groups of 6-7 children. Participants attended an average of 16.4 sessions.	RCT with 5 conditions	WLC: 26 CT + Teacher Training (TT): 23 PT + TT: 24 PT: 31 PT + CT + TT: 25	Mother Reported CBCL - Internalising	Findings for the intervention compared to waitlist control were non-significant, with no effect. There was also no significant change in the combined CT + TT condition. In a two-group contrast a significant contrast was found for the PT + CT + TT condition against the control group (p = 0.006). This combined condition also showed a significant improvement over just the PT condition (P=0.04). A secondary analysis was undertaken on the participants with elevated internalising symptoms at baseline, this showed a two group contrast of -7.27 with a p value of 0.07.

#5 Webster Stratton, Reid & Hammond, 2004					
Participants	**Intervention**	**Design**	**Comparison Group**	**Measures**	**Outcomes / Key Findings**
159 participants aged 4 to 8 years with a diagnosis of Oppositional Defiant Disorder.	30 children received the target intervention (CT) for 18 – 19 weeks in groups of 6-7 children. Participants attended an average of 16.4 sessions.	RCT with 5 conditions	WLC: 26 CT + Teacher Training (TT): 23 PT + TT: 24 PT: 31 PT + CT + TT: 25	Parent Positive and Negative Composite Score (using two parent reports and two observational measures) Child Conduct Problems – Home Composite Scores (combining the EBCI-Intensity scale, a parent report variable and behavioural observations using CII and DPICS-R) Child conduct problems at school and with peers composite score (using the TASB – Teacher Report Measure, PCSC – Teacher Rating Scale, MOOSES classroom observation coding system, SHP – a	The results showed a significant improvement in all child composite score measures for the intervention condition compared to control, apart from father reported child conduct problems at home. The father reported scores showed significant improvement in all conditions which included the Parent Training intervention. The social competence with peers composite score showed no significant change in the PT vs WLC or PT + TT vs WLC but did show a significant improvement in the target intervention condition at a $p<.05$ level. Where the CT intervention was combined with TT significant

				revised version of the Teacher Observation of Classroom Adaptation and the DPIS Coding system) Child Social Competence with peers Composite Score (using TASB – Teacher Report Measure, SHP and DPIS scores) Negative Classroom Management and Atmosphere Composite Score (using MOOSES, three items from the Teacher Coder Impression Inventory and an observation of classroom atmosphere)	changes were found in the mother reported conduct at home scores and conduct at school scores. In the combined CT + PT +TT condition all measures showed a significant improvement to $p<0.01$.

#6 Hutchings et al. 2007					
Participants	**Intervention**	**Design**	**Comparison Group**	**Measures**	**Outcomes / Key Findings**
9 children aged 7 – 11. 8 boys and 1 girl, 5 had	17 weeks (1 week less than recommended	Single case design	N/A – Single case design	Strengths and Difficulties Questionnaire (SDQ) – Parent and Teacher	Significant improvements were found on three parent report measures of child behaviour, the

diagnosis of conduct disorder and 4 had diagnosis of conduct disorder with ADHD. 3 in mainstream school, 4 attending pupil referral units and 1 attending a specialist school and 1 was excluded from school.	minimum). 2 hour weekly sessions following the dinosaur school curriculum. Attendance was not reported.		completion. SDQ and SDQ Impact scores are reported. Self-Control Rating Scale (SCRS) – Parent and Teacher completion Eyberg Child Behaviour Inventory (ECBI) – Parent Completion. Measures of Intensity and Problems are reported	ECBI – Intensity, the SDQ impact score and the SCRS. Teacher reports of the SDQ and SCRS both showed significant improvements. Post-intervention results were compared against a clinical cut off value. Cohen's effect sizes (which were not quoted) were then used to judge 'clinically significant improvements' which are quoted as being achieved for all measures despite clinical cut-off values being not being achieved. Calculated effect sizes via Lipsey and Wilson (2001) were large.

Appendix 3: Weight of Evidence Criteria

WoE A. Methodological Quality (Based on Kratochwill's coding protocol and Horner's Quality Indicators)

Measures

Weighting	Description
High (3)	(At least 3 out of 4) • Reported reliability of .85 or above for all primary outcomes relevant to the review question • Uses multiple methods of assessment (more than one) • Uses multiple sources of assessment (more than one) • Measures used are standardised or norm-referenced for the target population, or high validity is reported.
Medium (2)	(At least 2) • Reported reliability of .70 of above for all primary outcomes relevant to the review question • Uses multiple methods of assessment (more than one) • Uses multiple sources of assessment (more than one)
Low (1)	(At least 1) • Reported reliability of .50 or above for primary outcomes for some primary outcomes • Uses well referenced, standardised or norm-referenced measures • Uses more than one method of assessment • Uses more than one source of assessment

Either 1) Comparison Group

Weighting	Description
High (3)	(All must be met) • An active comparison group is used • Group equivalency is established • Evidence is given of equivalent mortality and low attrition
Medium (2)	(All must be met) • A “no intervention” comparison group is used but no active comparison group (e.g. Waitlist control) • Group equivalency is established • Evidence is given of equivalent mortality and low attrition
Low (1)	• A comparison group is used • One of the following criteria is met: (1) group equivalency is established (2) evidence is given of equivalent mortality and low attrition (3) evidence is given that change agents were counterbalanced

Or 2) Single Case Design

Weighting	Description (All criteria within each category must be met)
High (3)	• The study used a multiple baseline design. • There were at least three demonstrations of intervention effect.

	• The study included generalisation and maintenance phase with at least three data points for each. • The participants, and process for selecting them, is described with sufficient detail to be replicated. • The dependent variables and measures are described with sufficient precision for replication.
Medium (2)	• The study made three attempts to demonstrate intervention effect as well as either generalisation or maintenance phases with at least three data points for each and sufficient detail to replicate the dependent variables and measures.
Low (1)	• The study may demonstrate intervention effect less than three times. Generalisation or maintenance data may not be included, or there may be less than three data points for each. There may not be sufficient detail about the DV and measures to replicate.

Analysis

Weighting	Description
High (3)	(All four) • The unit of analysis is appropriate • Familywise/experimenter wise error is controlled (when applicable) • The sample size (N) is sufficiently large in all groups

	• Pre and Post Measures are reported and effect sizes are given (or enough information is given to calculate effect sizes) for all outcomes
Medium (2)	(At least 2) • The unit of analysis is appropriate • Effect sizes are calculated or enough information is given to calculate effect sizes for most outcomes • Familywise/experimenter wise error is controlled (when applicable) • Pre and post measures are reported • The sample size (N) is sufficiently large in all groups
Low (1)	(At least 1) • The unit of analysis is appropriate • Enough information is given to calculate some effect sizes • The sample size (N) is sufficiently large in all groups • Familywise/experimenter wise error is controlled (when applicable) • Pre and post measures are reported

(In this category the criteria relating to group sample size does not apply to single case design studies)

In order to ascertain an overall score for WoE A each area was assigned a score of 3, 2 or 1 and an average score was calculated: **(Measures + Comparison + Analysis) ÷ 3 = WoE A**

B. Methodological Relevance

Either 1) Comparison Group

Weighting	**Description** (All criteria within each category must be met)
High (3)	• The study must have used an active comparison group and “no intervention” group (waitlist control). • The participants must have been randomly assigned to either group and group equivalences must be demonstrated. • Quantitative pre and post-test measures must have been conducted for all groups and all outcomes. • There must be more than one method undertaken to ensure fidelity to the intervention.
Medium (2)	• The study must have used a comparison group (either alternative intervention or waitlist control). • Group equivalences must be demonstrated and pre and post-test measures must have been conducted for all groups and outcomes. • There is evidence that one criteria has been met to ensure fidelity to the intervention.
Low (1)	• At least one of the above criteria must have been met

Or 2) Single Case Design

Weighting	Description
High (3)	• The study must have made use of multiple baseline design and included at least three demonstrations of intervention effect. The study must include generalisation and maintenance phases with at least three data points for each.
Medium (2)	• The study must have made three attempts to demonstrate intervention effect, as well as either generalisation or maintenance phases with at least three data points for each.
Low (1)	• The study may demonstrate intervention effect less than three times. Generalisation or maintenance data may not be included, or there may be less than three data points for each. There may not be sufficient detail about the DV and measures to replicate.

C. Topic Relevance

Weighting	**Description (**All criteria within each category must be met)
High (3)	• Dependent variable included measures of emotional and behavioural outcomes • The Intervention group received only the Therapeutic Dinosaur School Programme. • The Intervention takes place in a school or setting environment.

	• A high level of fidelity to the intervention was evidenced (multiple criteria for fidelity were met and the intervention was a minimum of 18 weeks long) • Demographics of the participants are described with detail relating to their diagnosis of Special Educational Need.
Medium (2)	• Dependent variable included measures of emotional and behavioural outcomes • The Intervention group received only the Therapeutic Dinosaur School Programme. • The level of fidelity to the intervention was reported, any amendments were described in detail with justification.
Low (1)	• Emotional or behavioural outcomes were measured • The level of fidelity to the intervention was reported. • The intervention group received multiple levels of the Incredible Years Programmes (Therapeutic Dinosaur School Programme and either Parent Programme or Teacher programme)

D. Overall Weight of Evidence

(WoE A + WoE B + WoE C) ÷ 3 = WoE D

WoE Scores were rated as follows:

1. Low < 1.4
2. Medium 1.5 – 2.4
3. High 2.5 +

Weight of Evidence Scores:

Study	WoE A	WoE B	WoE C	WoE D (Overall weight of evidence)
7. Larsson et al., 2009	*2.3*	*3*	*1*	*2.1 (medium)*
8. Webster-Stratton, Reid & Beauchaine, 2011	*2*	*2*	*1*	*1.7 (medium)*
9. Webster-Stratton & Hammond, 1997	*2*	*3*	*2*	*2.3 (medium)*
10. Herman et al. 2011	*2*	*2*	*1*	*1.7 (medium)*
11. Webster Stratton, Reid & Hammond, 2004	*2.7*	*3*	*2*	*2.6 (high)*
12. Hutchings et al. 2007	*1.3*	*1*	*2*	*1.4 (low)*

Appendix 4: Group Coding Protocol

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

Group Design Coding Protocol

Name of Coder: X Date: 23.01.2017

Full Study Reference in proper format: Larsson, B., Fossum, S., Clifford, G., Drugli, M. B., Handegård, B. H., & Mørch, W.-T. (2009). Treatment of oppositional defiant and conduct problems in young Norwegian children. *European Child {&} Adolescent Psychiatry*, *18*(1), 42–52.

Intervention Name (description of study):_The Incredible Years Dinosaur School Programme and Parent Programme

Study ID Number:_#1 _

☐ Type of Publication:

☐ Book/Monograph

☒ Journal Article

☐ Book Chapter

☐ Other (specify):

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. Participants

Total size of sample (start of study): __136____

Intervention group sample size:_55_(PT + CT)

Control group sample size:_30_(WLC) 51 (PT)

PT = Parent Training, CT = Child Training (Dinosaur School Programme)

C. Type of Program

☐ Universal prevention program

☐ Selective prevention program

☐ Targeted prevention program

☒ Intervention/Treatment

☐ Unknown

D. Stage of Program

☐ Model/demonstration programs

☒ Early stage programs

☐ Established/institutionalized programs

☐ Unknown

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 (mother and father report)

☐ 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_2_

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

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 (not compared in this review, 3 group conditions in total)

☐ 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_

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

C. Appropriate Statistical Analysis

Analysis 1_____

☒ Appropriate unit of analysis

☒ Familywise/experimenter wise error rate controlled when applicable

☐ Sufficiently large N

Analysis 2_____

☐ Appropriate unit of analysis

☐ Familywise/experimenter wise error rate controlled when applicable

☐ Sufficiently large N

Analysis 3_____

☐ Appropriate unit of analysis

☐ Familywise/experimenter wise error rate controlled when applicable

☐ Sufficiently large N

Overall rating for Statistical Analysis _2___

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

D. Fidelity of intervention

Evidence of fidelity is reported through:

☒ Multiple examples are reported

☒ Ongoing supervision/consultation/audit

☒ Formal training in intervention delivery

☐ Coding Sessions

☐ Audio/video tapes

☒ Use of a manual

☐ No evidence is reported.

Overall rating for fidelity ___3_____

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

Appendix 5: Single Case Design Coding Protocol: Quality Indicators (Horner 2005)

Name of coder: X
Date: 23.01.2018
Study number: 6

Full study Reference in APA Format: Hutchings, J., Bywater, T., Daley, D., & Lane, E. (2007). A Pilot Study of the Webster-Stratton Incredible Years Therapeutic Dinosaur School Programme. *Clinical Psychology Forum*, *170*, 21–24.

	Is the Criteria Met?	
a. Description of Participants and setting		
	Yes	**No**
• Participants are described with sufficient detail to allow others to select individuals with similar characteristics (e.g., age, gender, disability, diagnosis)	Yes	
• The process for selecting participants is described with sufficient detail to allow for replication		No
• Critical features of the physical setting are described with sufficient precision to allow replication		No

	Is the Criteria Met?	
b. Dependent Variable (DV)		
	Yes	**No**
• Dependent variables are operationalised and described with sufficient precision to be replicable	Yes	
• Each dependent variable is measured with a procedure than generates a quantifiable index	Yes	
• Measurements are valid and described with sufficient precision to be replicated	Yes	

• Dependent variables are measured repeatedly over time		No
• Data is reported on reliability of each dependent variable and IOA levels meet minimal standards (e.g.,IOA = 80%; Kappa = 60%).		No

	Is the Criteria Met?	
c. Independent Variable (IV)		
	Yes	**No**
• IV is described with sufficient precision to be replicable		No
• IV is systematically manipulated and controlled by the experimenter		No
• Overt measurement of the fidelity of implementation for the independent variable is highly desirable		No

	Is the Criteria Met?	
d. Baseline		
	Yes	**No**
• Baseline conditions are described with sufficient precision to be replicated	Yes	
• The baseline phase provides repeated measurement of the DV to establish a pattern of responding that could be used to predict the pattern of future performance, if introduction or manipulation of the independent variable did not occur.		No

	Is the Criteria Met?	
e. Experimental Control/ Internal Validity		
	Yes	**No**
• The design provides at least three demonstrations of experimental effect at three different points in time.		No

• The design controls for common threats to internal validity (e.g., permits the elimination of rival hypotheses).		No
• The results document a pattern that demonstrates experimental control		No

	Is the Criteria Met?	
f. External Validity		
	Yes	**No**
• Experimental effects are replicated across participants, settings, or materials to establish external validity.	Yes	

	Is the Criteria Met?	
g. Social Validity		
	Yes	**No**
• The dependent variable is socially important	Yes	
• The magnitude of change in the dependent variable resulting from the intervention is socially important		No
• Implementation of the independent variable is practical and cost effective.		No
• Social validity is enhanced by implementation of the independent variable over extended time periods, by typical intervention agents, in typical physical and social contexts.		No

A: 1 **B**: 3 **C**: 0 **D**: 1 **E**: 0 **F**: 1 **G**: 1