

Case Study 1 – An Evidence-Based Practice Review Report

Does the “Tuning into Kids” emotion coaching parental program support parents to reduce child behavioural difficulties and to what extent is the program relevant to current Educational Psychology practice in the UK?

1. Summary

This systematic literature review aims to evaluate the effectiveness of the parental emotion coaching intervention, “Tuning in to Kids”. The program aims to not only develop parental knowledge and skill but also to demonstrate a reduction in childhood behavioural difficulties and it is the latter that is the focus of this review. Six studies that met the inclusion criteria were included and a combination of both small and large statistically significant effect sizes on child behavioural outcomes were found. This indicates that the program is somewhat effective at reducing child externalising behaviour in participants with and without existing behavioural difficulties. As indicated by the studies showing the larger effect sizes, the program seemed to be most effective when utilised as a targeted approach for children at risk of conduct disorders. Potential use and implications for future directions are discussed.

2. Introduction

2.1. Emotional competence and the role of parents

Denham, Bassett, and Wyatt (2007) posit that 'emotional competence' (the ability to express, understand and appropriately regulate emotions) develops through early childhood and into adolescence through a combination of individual factors and interactions with others (Denham et al., 2007; Eisenberg, Cumberland, & Spinrad, 1998). Through the lens of both Social Learning Theory (Bandura, 1977) and Vygotskian ideas (Vygotsky, 1978), the development of emotional competence is supported through parental socialisation of emotion. In summary, this relates to the way in which parents and significant others model emotional understanding and regulation of their own emotional experiences, react to a child's emotions and the way in which they teach a child about emotions and how to manage these (Denham et al., 2007). Research findings suggest that parents who respond more supportively to negative child emotions are likely to have more accepting beliefs about children's negative emotion (Wong, McElwain & Halberstadt, 2009) and that emotional supportiveness in parents has a direct contribution to emotional regulation in toddlers (Bocknek, Brophy-Herb, & Banerjee, 2009). Parenting styles and emotional expressiveness therefore impact upon family emotional climate and child emotional development (Morris, Silk, Steinberg, Myers, & Robinson 2007).

2.2. Emotion competence and a meta-emotion philosophy

Gottman, Katz, and Hooven (1996) describe the notion that parents have established thoughts, feelings and beliefs about theirs and their children's

emotions termed 'parental meta-emotion philosophy' (PMEP). It is argued that PMEP is influential in child emotion socialisation in that it shapes parental responses to emotions in children and consequently the emotional messages they receive (Katz, Maliken, & Stettler, 2012). Of these, Gottman and colleagues (1996) describe a favourable PMEP termed 'emotion-coaching' (see Table 1).

Table 1.

The five elements of an emotion coaching PMEP.

1	Noticing the emotions, especially emotions of lower intensity such as frustration.
2	Viewing emotions as an opportunity for intimacy and teaching.
3	Showing an understanding and acceptance of a child's emotions, positive or negative.
4	Supporting a child to name and label the emotions that they experience.
5	Assisting in problem solving and where necessary reinforcing boundaries.

They discriminated this from other less helpful approaches including a dismissive approach (plays down, ridicules or uses distraction), a disapproving approach (harsh, judgmental and critical) or a laissez faire approach (permissive, does not set boundaries, offers little guidance).

Research suggests that PMEP explains the variance in a child's socioemotional adjustment that cannot be simply accounted for by parental warmth or harshness (Sheeber, Shortt, Low & Katz, 2010). Research also

supports an association between PMP and attachment theory (Bowlby 1973), for example children whose mothers tended to adopt an emotion coaching PMP were more likely to experience secure attachments whereas a dismissing PMP was related to lower levels of attachment security (Chen, Lin & Li, 2012). Children with secure attachment styles are frequently reported to demonstrate more effective coping skills and are found to be more socially competent (Contreras, Kerns, Weimer, Gentzler & Tomich, 2000). PMP may therefore act as a means of supporting strong attachments and the development of both emotional and social competence.

2.3. Emotion coaching and externalising behaviour problems

Given likely variations in both individual and parenting factors, not all children appropriately acquire the emotional competence required to successfully manage emotional and social situations. However this leaves these children at considerable risk of a whole host of negative outcomes including current and future problematic behaviour and psychopathology (Denham et al., 2007), especially when there has been less focus on the socialisation of negative emotions (Johnson, Hawes, Eisenberg, Kohlhoff & Dudeney, 2017). Correlational evidence supports the association between a supportive, emotion coaching (as opposed to dismissing) approach, better child emotional regulation and subsequently reduced disruptive and challenging behaviours in their children (Duncombe, Havighurst, Holland & Frankling, 2012; Dunsmore, Booker & Ollendick, 2013; Loop & Roskam, 2016). It therefore seems pertinent that efforts are made to support successful emotional competence through teaching parents emotion coaching skills.

2.4. A parent emotion coaching intervention: Tuning in to Kids

Program rationale and aims

“Emotion Coaching” (EC), as outlined by Gottman and DeClaire (1997), is a strategy that supports a positive relationship between children and key adults aiming to enable children to develop emotional competence and reduce inappropriate or destructive behaviour (Gus, Rose & Gilbert, 2015; Shortt, Stoolmiller, Smith-Shine, Eddy, & Sheeber, 2010). The “Tuning in to Kids” (TIK) parental emotion coaching program was developed by Havighurst, Wilson, Harley, and Prior (2009) with these aims in mind. Originally for parents with preschool age children, it has since been applied across the age range from 18 months (Tuning in to Toddlers or TOTS) to early adolescence (Tuning in to Teens or TINT) and a slightly longer father’s specific version (Dads TIK). For simplicity, the review will refer to the intervention as “TIK” throughout unless there is a reason to refer to one of the variants specifically.

Program content

TIK is a manualised program that involves six two-hour weekly sessions facilitated by a trained TIK facilitator. The program teaches parents the five steps of emotion coaching (Gottman & DeClaire, 1997) in a gradual and incremental process across the first five sessions with the last of the sessions acting to consolidate the five steps and providing an opportunity for parents to consider how they will apply the learning and skills in the future. Table 2 illustrates the main goals of each session with a brief description of session content (Havighurst, Kehoe, Harley, & Wilson, 2015, page 44).

Table 2.

Tuning in to Kids goals and program content

Session	Goals	Content
1	To engage parents; to validate and normalise parents' experiences; to introduce the concept and skills of emotion coaching	Normalising parents' parenting experience What is emotional intelligence and why is it important? 5 steps of emotion coaching Noticing emotions at a lower intensity
2	To increase parents' emotion awareness and recognition; examine beliefs about emotions; explore emotion dismissing versus emotion coaching	Emotion awareness and recognition. Emotion Tuning Meta-emotion philosophy Parenting styles (Dismissing and Coaching) Labelling emotions
3	To build empathy; develop emotion vocabulary; build and refine emotion coaching	Mindfulness meditation The emotion detective to develop empathy Role-play practices of emotion coaching
4	To explore effects of criticism; consolidate and extend emotion coaching; explore emotion coaching with sadness/worry; increase parent emotional self-care	Mindfulness meditation Self-care Responding to fears and worries Problem solving
5	To increase understanding of anger; build skills in managing parent's and child's anger; understand the need for boundaries and negotiation around behaviour	Understanding and responding to anger in parent and child Managing sibling conflict
6	To consolidate the emotion coaching skills; increase support and self-care	Role plays Review principles of emotion coaching

2.5. Rationale and relevance

Children and young people with lower levels of emotional competence are considered more at risk for, amongst other things, disengagement from school and poorer levels of academic attainment (Gutman & Vorhaus, 2012). Prevention and early intervention work in supporting these at risk children is an area of work that is recommended by and increasingly adopted in practice by Educational Psychologists (Farrell et al., 2006 and Noble & Mcgrath, 2012). McGuiggan (2017) reported that EPs expressed a desire to work in a more eco-systemic way with families. They cited being community based and having the appropriate knowledge and skill base with which to support children and families (McGuiggan, 2017). Stringer, Powell, and Burton (2006) supported this citing numerous ways in which EPs can engage in community psychology activities. The TIK program therefore has the potential to add to the repertoire of the systemically minded Educational Psychology Service for supporting parents in being able to prevent or reduce behavioural difficulties in their children.

2.6. Review question

Does the “Tuning into Kids” emotion coaching parental program support parents to reduce child behavioural difficulties and to what extent is the program relevant to current EP practice in the UK?

3. Critical Review of the Evidence Base

3.1. Literature search

Initial searches were carried out in December 2018 using electronic databases; Web of Science, ERIC, Child Development and Adolescent Studies (CDAS), PsychINFO and other scoping searches. A multi-field search within 'all fields' was conducted in each of these databases using search terms relevant to the research question, see Tables 3a and 3b below. The key word 'tuning' was omitted from the latter three databases as this narrowed the search results significantly. By removing this word some important articles were found and included for review.

Table 3a

Search terms used within the Web of Science database

Search terms	Field
Tuning	ALL
AND child* or kid* or preschooler* or toddler* or tot* or teen* or adolescen* or youth or young person	ALL
AND intervention or training or program* or approach or parenting	ALL
AND emotion* and coaching or socialisation or socialization or focus* or competence	ALL
AND behavio* or conduct or externalizing or difficult* or problems	ALL

Table 3b

Search terms used for ERIC, Child Development and Adolescent Studies (CDAS) and PsychINFO databases

Search terms	Field
emotion*	
AND coaching or focused or socialization	ALL
AND intervention or training or program or group or approach	ALL
AND parenting or parental	ALL
AND behavio* or conduct or externalizing or difficult* or problems	ALL
AND child* or kid* or preschooler* or toddler* or tot* or teen* or adolescen* or youth or young person	ALL

* = wildcard search term

Whilst conducting the scoping searches, an unpublished systematic literature review paper of the TIK program was found dated December 2014 (Osei, 2015). Of the six articles reviewed by Osei, one was included within the current review as it was conducted within the last five years. By focusing solely on research within this time frame, this review provides an updated and original analysis of the effectiveness of the TIK program.

3.2. Inclusion and exclusion criteria

Studies retrieved were included in the review if they met the criteria as detailed within Table 4. Figure 1 illustrates that a total of 568 articles were found through electronic database searches and other scoping searches. Of

these, 46 were removed as duplicates with the remaining 522 being screened by title. Screening by title highlighted a large number of articles of no or very little relevance to the research question (446) after which 76 were screened at the abstract level. Abstract screening revealed 57 papers that did not meet the inclusion criteria and as a result, 19 articles were screened at the full text level. Of these thirteen were excluded from the review using the criteria in Table 4 (see Appendix 1) and six studies were included (see Appendix 2).

Table 4.

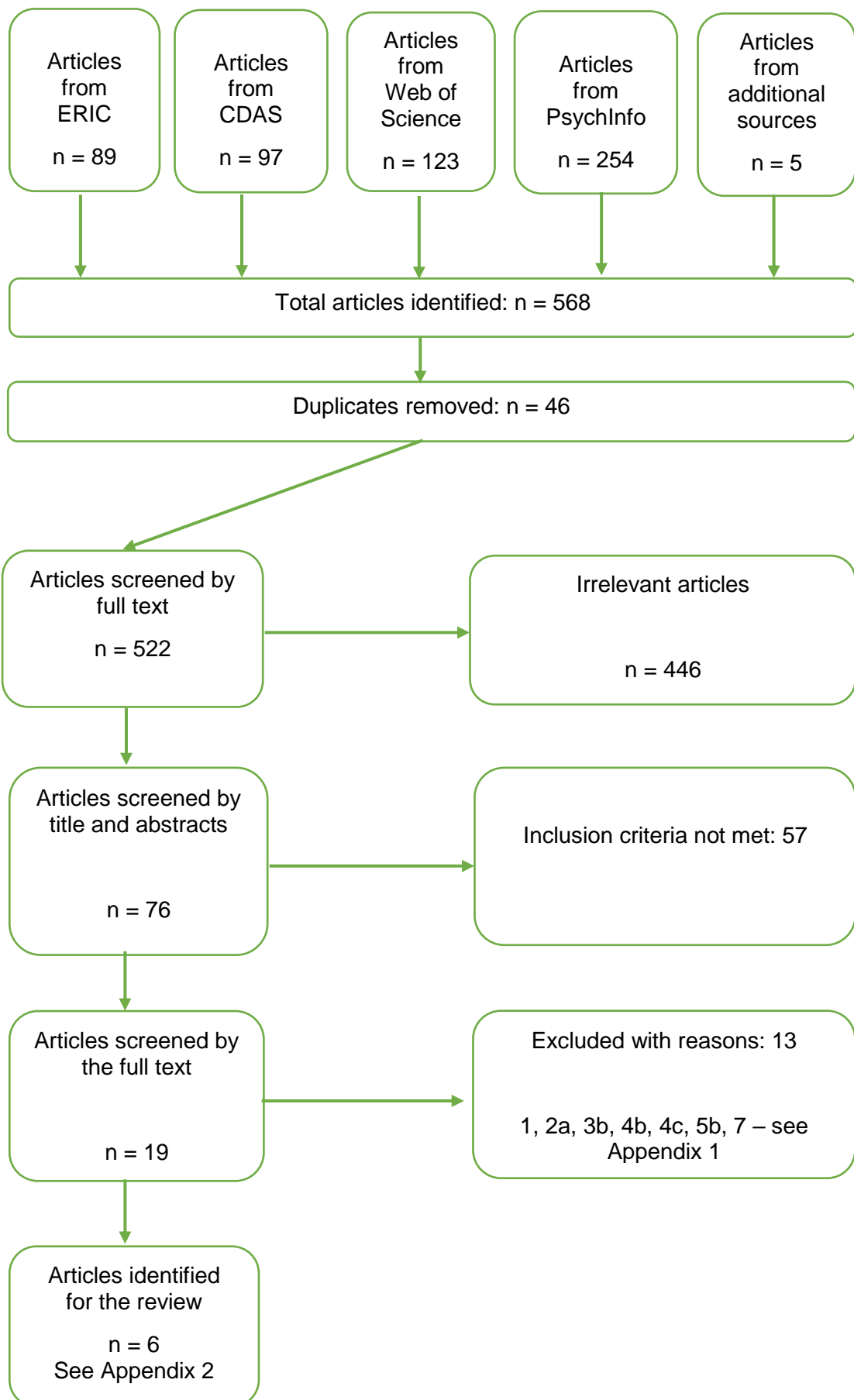
Inclusion and exclusion criteria

Criteria	Inclusion	Exclusion	Rationale
1. Type of publication	Must be in a peer-reviewed journal or have been submitted for review.	Is not in or has not been submitted for a peer-reviewed journal, i.e. book chapters, conference papers, unpublished theses.	Peer reviewers assess the quality of research articles and so these are likely to meet the required standards.
2. Design	<ul style="list-style-type: none"> a) Uses primary, empirical, quantitative data. b) Assesses effectiveness through an experimental or quasi-experimental design. 	<ul style="list-style-type: none"> a) Qualitative design or design that does not use original data. b) Design method that fails to use an experimental or quasi-experimental design. 	This review question seeks to assess the effectiveness of an intervention and thus requires an experimental design.
3. Intervention	Use of the ' <i>Tuning in to Kids</i> ' (or other adapted versions) emotion coaching intervention, delivered to parents for the purpose of improving emotional regulation/reducing behavioural problems in their children.	<ul style="list-style-type: none"> a) Emotion coaching training involving other professionals such as school staff b) A parental emotion coaching training program that isn't '<i>Tuning into Kids</i>' or its variants. 	This review question aims to consider the effectiveness of the ' <i>Tuning in to Kids</i> ' parental emotion coaching program only. Other/non-specific emotion coaching interventions and those not delivered to support parents and their children fall outside the scope of this review.

Criteria	Inclusion	Exclusion	Rationale
4. Measures and outcomes	<ul style="list-style-type: none"> a) Quantifiable measure(s) of effectiveness are evident. b) Measures administered specifically assess emotion management and behavioural conduct (externalising behaviours). c) Outcomes reported must include outcomes for children but outcomes for parents can also be reported. d) Measures can be administered to parents, teachers, children and young people, as well as being observation or direct assessment based. 	<ul style="list-style-type: none"> a) No quantifiable measures of effectiveness in relation to a child's emotional regulation are used. b) Measures focus on child/young person internalising behaviours or knowledge gain with no additional measures of externalising behaviours present. c) Only parental outcomes are measured and reported. 	<p>An important feature of the review question is the impact the intervention has on children, thereby requiring studies to include a quantifiable measure of child behavioural difficulties (alternative wording: externalising behaviour, behaviour problems and conduct difficulties).</p>
5. Participants	<ul style="list-style-type: none"> a) Parents (including foster carers or adoptive parents) of children with emotional dysregulation, behavioural or conduct difficulties. 	<ul style="list-style-type: none"> a) Any studies where the intervention solely took place within an education setting with school staff. 	<p>This review focuses on the impact of the '<i>Tuning in to Kids</i>' emotion coaching parent program on preschool and school age children as preschool age and onwards is considered a key time</p>

Criteria	Inclusion	Exclusion	Rationale
	<ul style="list-style-type: none"> b) Parents of pre-school and school age children aged over 3 years of age. c) Children with no additional SEND needs e.g. autism or where they have coincidental additional needs, these are not the focus of the intervention. 	<ul style="list-style-type: none"> b) Parents of children aged 0-2 years and young people aged 19+. c) Studies where additional SEND needs (e.g. children with ASD) are the priority focus of the intervention. 	<p>point for the development of emotional competence and also the presence of language at this age would be supportive of such an intervention.</p>
6. Language and setting	Must be published in English but can be undertaken in any country.	Any publications that are not written in English or do not have the English translation already provided.	The reviewer is unable to speak other languages and so would not be able to understand and evaluate the findings. Using online translation services may not produce a reliable translation.
7. Date of publication	Studies published from 2013 onwards.	Studies published prior to 2013.	The review intends to consider the most current research evidence, i.e. published in the last 5 years.

Figure 1: flowchart showing application of inclusion and exclusion criteria



3.3. Critical appraisal for quality and relevance

The six included studies were evaluated for their research quality and relevance using Gough (2007)'s weight of evidence (WoE) framework. This allowed for a systematic critical appraisal of all the research and comparisons between the studies to be drawn. The procedure involves evaluating the research on four aspects: methodological quality (WoE A), methodological relevance in relation to the review question (WoE B), the relevance of the focus of the study to the current review question (WoE C) and an overall weight of evidence (WoE D). Kratochwill (2003)'s coding protocol for group-based design was used and adapted for the purpose of this review to assess WoE A (see Appendix 4c). All studies were coded using the same protocol to maintain consistency however a rating on 'comparison group' was omitted from the one within groups design study. A table detailing the rationale for the adaptations, can be found in Appendix 4a and an example of a completed coding protocol can be found in Appendix 4b. WoE B was supported through the use of Guyatt et al.'s (1995) hierarchy of evidence with criteria for assessing this aspect detailed in Appendix 5. Table 5 provides a summary of the WoE ratings across all of the six studies.

3.4. Study design

Comparison group

Only studies of an experimental or quasi experimental design were included within the review. All but one of these (Wilson, Havighurst, & Harley, 2014) were between groups designs utilising some form of comparison group which enables the effects identified to be more robustly attributed to the

Table 5.

WoE table

Studies	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. Havighurst, Duncombe et al. (2015)	Medium (2.00)	Medium (2.00)	High (3.00)	Medium (2.30)
2. Havighurst, Kehoe, and Harley (2015)	Medium (1.75)	Medium (2.00)	Medium (2.00)	Medium (1.90)
3. Havighurst et al. (2013)	High (2.50)	Medium (2.00)	High (3.00)	High (2.50)
4. Meybodi et al. (2017)	Medium (2.25)	Medium (2.00)	Medium (2.00)	Medium (2.10)
5. Wilson, Havighurst, and Harley (2014)	Low (1.00)	Low (1.00)	Low (1.00)	Low (1.00)
6. Wilson et al. (2014)	Medium (1.75)	Medium (2.00)	Low (1.00)	Medium (1.60)

* < 1.5 = low; 1.5 – 2.4 = medium; > 2.4 = high

intervention. One study utilised an 'active' comparison group as opposed to a waiting list control group where the control group received 'treatment as normal' (Havighurst et al., 2013). This is favourable as there is no delay in the control group receiving some kind of support and it allows for conclusions to be drawn about the intervention being more effective than other interventions. As a result, this study was rated high on WoE A and B.

Wilson et al. (2014) used a within participant pretest-posttest design with no control group. As the intervention in question relates to outcomes for

children, it is possible that the results could be confounded by maturation effects for example (Cook and Campbell cited in Barker, Pistrang & Elliot, 2002). Additionally, one study (Havighurst & Duncombe et al., 2015) evaluated TIK as part of a multi-systemic approach with additional teacher and child interventions running concurrently without any attempts made to analyse the separate parts of the intervention. In both of these cases it is more difficult to establish whether it was TIK or other factors that accounted for the changes reported thus contributing to lower WoE B scores.

Participants

The selection of participants is highly relevant to WoE C considerations and these will be discussed more fully within this section whilst participant characteristics are described more fully within the Mapping the Field table (Appendix 3). WoE C criteria (see Appendix 6, Table 1) were focused on the relevance to the review question and thus the comparability of the research to the typical practice of an educational psychologist within a UK context was important. Havighurst and Duncombe et al. (2015) and Havighurst et al. (2013) achieved high WoE C ratings, with the former meeting all the criteria and the latter meeting all but one; the children were drawn from a clinical population which may not be within the remit of an Educational Psychologist. Meybodi, Mohammadkhani, Pourshahbaz, Dolatshahi, and Havighurst (2017) and Havighurst, Kehoe, and Harley (2015) were rated medium on WoE C with the findings of the former being less generalizable as the participants were drawn from an area of high socio economic status in Iran and the latter

were also largely of higher socio-economic status and children were not screened for possible behavioural difficulties.

The two lowest rated studies in terms of participants as reflected in low WoE C ratings were Wilson et al. (2014) and Wilson, Havighurst, Kehoe, and Harley (2016). In these studies, parents selected were solely fathers which is atypical of parental involvement with EPs. In addition, a large proportion of the families had high household income and fathers were self-selected without any screening for those children who may be at risk for behavioural difficulties. This method of convenience sampling introduces an element of bias into the sample as not every member of the target population has an equal chance of being selected (Barker et al., 2002). It is possible that characteristics of participants who self-select, for example motivation to improve parenting, an interest in understanding and supporting child emotional development, may differ to those who fail to do so thus limiting the extent to which findings can be generalised.

Randomisation of participants

Havighurst et al. (2013) and Meybodi et al. (2017) randomised participants at the individual level which allows for bias to be more effectively controlled for, thus reducing the impact of confounding variables and allowing for more accurate estimation of effect. This contributed towards their WoE B scores. Havighurst and Duncombe et al. (2015), Havighurst, Kehoe, and Harley (2015) and Wilson et al. (2016) randomised participants at the group level by school or preschool. On the one hand, this can be regarded as beneficial in

that it minimises contamination of the treatment to the control group. It could be argued however that as this is an intervention with parents and not with school staff, the treatment contamination may in actuality have been negligible. Randomisation by setting introduces a potential for bias in that it may be difficult to determine whether the effects are attributable to the intervention or some characteristic of the particular school. Positively, all three studies recognised the possibility for confounding effects by school and accounted for any group differences in post-hoc analyses.

Measures

Of all the studies Havighurst and Duncombe et al. (2015) and Havighurst et al. (2013) were notable for the use of multiple methods of evaluating and multiple sources, supporting a higher WoE A score for these studies on this component. The other studies within this review utilised a variety of outcome measures with one adding an additional source other than parents (Havighurst, Kehoe and Harley, 2015) whilst the others relied solely on parental report measures. All studies within this review used measures with established reliability and validity and also reported acceptable or high levels of internal consistency reliability. The latter is important as effects are more likely to be attributable to the intervention rather than some variation caused by the measurement tool.

It should be noted that outcome measures were largely focused on parental outcomes with much less attention given to child outcomes. One WoE C criterion stipulated that child outcomes were either primary outcomes or at

least two methods or two sources were used to measure child outcomes. Three of the studies achieved this criterion; Havighurst and Duncombe et al. (2015), Havighurst, Kehoe, and Harley (2015) and Havighurst et al. (2013). The remaining three studies used only one child outcome measure, reported on by only one source (parents) and so this contributed to their capacity to achieve a higher WoE C score.

Program implementation

A common theme across the research studies was a clear commitment to implementation fidelity with the majority describing in detail how high fidelity had been achieved e.g. through experienced facilitators, detailed manuals and fidelity checks. Wilson et al. (2014) provided insufficient detail to ascertain the level of rigour applied to implementation fidelity and this contributed to a lower WoE A score.

Follow up

The studies that included a follow up evaluation (Havighurst et al., 2013 and Meybodi et al., 2017) scored higher on WoE A and B than those that only used two time points as this allowed for a more robust assessment of whether the intervention effects are maintained after participants are no longer receiving the treatment. Two studies; Havighurst and Duncombe et al. (2015) and Havighurst, Kehoe, and Harley (2015) completed their post intervention outcome measures 10.5 and 10 months later respectively rather than immediately after. The absence of immediate post data means that it cannot be concluded with any certainty that it was the intervention, and not

some other factors that occurred in the time between, that were responsible for the effects observed. One of these studies, (Havighurst & Duncombe et al., 2015) had a high attrition rate and it is possible that the extended time delay between the intervention ending and the collection of post intervention data may have contributed to this.

Findings

There was a range of effect sizes demonstrated across the studies on parent reported child outcomes of behaviour (see Table 6). Due to the inconsistency between studies, this review will focus on pre and post effect sizes only. One study (Meybodi et al. 2017) included a three month follow up period and reported effect sizes as an interaction between condition and time. In this instance only, effect sizes discussed relate to pre and follow up scores. Three of the studies, Havighurst and Duncombe et al. (2015), Havighurst, Kehoe and, Harley (2015) and Wilson et al. (2016) reported statistically significant findings with small or small-medium effect sizes. Across these studies greater effect sizes were reported for parent outcomes perhaps indicating that the program has, in the short term at least, a greater impact on parents. However it was outside the scope of this review to consider parent outcomes in any depth. In addition, Havighurst and Duncombe et al. (2015) and Havighurst, Kehoe and, Harley (2015) conducted their post-intervention measures 10 and 10.5 months later which may be reflected in the lower effect sizes as treatment effects may be most pronounced immediately after the intervention.

The remaining three studies all reported statistically significant findings with large effect sizes. Of these, Havighurst et al. (2013) and Meybodi et al. (2017) screened participants for being at risk of behavioural difficulties and so it could be argued that the program is more effective as a targeted as opposed to preventative measure. Whilst regression to the mean might be a possible factor affecting the findings when 'high risk' participants are used (Linden, 2013), it is positive that both Havighurst et al. (2013) and Meybodi et al. (2017) randomised participants at the individual level and so both groups are likely to have been equally effected by this (Barnett, Van der Pols, & Dobson, 2005). All three studies evidencing large effect sizes used small samples. Some argue that small sample sizes are more variable, more influenced by sampling error, less representative of the population mean and thus produce less reliable and less replicable results than studies with larger sample sizes (Fan, 2001; Slavin & Smith, 2009). Therefore caution should be taken with regard to these results which may represent an overestimation of effect size or a greater likelihood that the results were confounded by sampling factors. Additionally, caution needs to be applied when interpreting Wilson et al. (2014)'s findings as this had a low overall WoE score.

Two of the studies (Havighurst et al., 2013 and Meybodi et al., 2017) considered the longevity of effects as measured by a follow up phase. Havighurst et al. (2013) reported that the interaction between time and condition for parent reported behaviour was greater in the first few months than in subsequent months but overall was found to be not significant. The significant positive effect sizes reported related to the post-hoc ANCOVA

comparing the two conditions at pre and post only. Using a repeated measures ANOVA, Meybodi et al. (2017) found a significant interaction between condition and time for child behaviour scores with a reduction for children in the intervention group at post-intervention that continued at follow up. Effect sizes relating to the difference between the pre and post scores were not calculated.

Two of the studies (Havighurst & Duncombe et al., 2015 and Havighurst et al., 2013) used parent and teacher reports of child behaviour and in these, the effect sizes found across both groups were very similar. It could be argued that a parental measure of child behaviour change could be influenced by expectancy effects bias as parents, knowing their children are part of a program designed to improve behaviour and having invested time and energy into learning and applying the new skills, may be motivated to report a reduction in post intervention problem behaviour. Therefore similar effect sizes reported by parents and teachers, teachers perhaps being less affected by such a bias, may be regarded as a strength of these two studies.

Table 6.

Summary of effect sizes and overall quality ratings (pre-post) for key outcome measures¹

Study	Source	Measure	Scale	Significance value (p) ²	Effect size (d) ³	Sample size	Overall WoE (D)
1. Havighurst and Duncombe et al. (2015)	Parent	The Eyberg Child Behaviour Inventory (ECBI)	Oppositional Defiance Disorder Conduct Disorder Hyperactivity	.011 .012 .066	.37 (small) .37 (small) .29 (small)	231	Medium
	Teacher	The Strengths and Difficulties Questionnaire (SDQ)	Total score	.003	.41 (small-medium)		
2. Havighurst, Kehoe and Harley (2015)	Parent	The SDQ	Conduct problems and hyperactivity-inattention	.003	.31 (small)	225	Medium
	Youth	The SDQ	As above	.030	.28 (small)		
3. Havighurst et al. (2013) <i>Pre to post</i>	Parents	The ECBI	Intensity	.009	.87 (large) ⁴	54	High
	Teacher	The Sutter-Eyberg Student	Intensity	.036	.87 (large) ⁴		
	Teacher	behaviour Inventory (S-ESBI)	Problems	.036	.87 (large) ⁴		

Study	Source	Measure	Scale	Significance value (p) ²	Effect size (d) ³	Sample size	Overall WoE (D)
4. Meybodi et al. (2017)	Parents (mothers)	The ECBI	Problem	.000	.92 (large) ⁴	54	Medium
<i>Pre to follow up</i>							
5. Wilson et al. (2014)	Parents (fathers)	The SDQ	Total	.010	.87(large) ⁴	43	Low
6. Wilson et al. (2016)	Parents (fathers)	The SDQ	Total	.024	.27(small)	162	Medium

¹ Only outcomes considered relevant to this review question are summarised here i.e. those that relate to child behavioural outcomes. Effect sizes were calculated from pre and post data with the exception of Meybodi et al. (2017) who reported effect sizes based on pre and follow up data.

² Significance levels: p<0.05 = significant, p<0.01 = highly significant

³ Effect size levels (Cohen's d): small = 0.2, medium = 0.5, large = 0.8.

⁴ Effect sizes originally reported as partial eta squared were converted to Cohen's d in order to be able to draw comparisons across the studies.

4. Conclusions

This systematic literature review aimed to evaluate the extent to which the TIK program supports parents to reduce behavioural difficulties in their children and to assess the relevance of the evidence to UK EP practice. The program appears to offer some universal benefits with at least small effects observed on child behavioural outcomes across all studies. Two of the three studies that reported large effects had screened participants through elevated scores on a behavioural measure and thus it could be argued that the program is the most beneficial when applied as a targeted approach with children at risk of behavioural difficulties. This supports conclusions drawn by Osei (2014) and also is congruent with current EP practice in the UK that typically prioritises intervention for children and young people with existing behavioural difficulties. Caution however does need to be applied when considering those studies with large effect sizes as the smaller sample sizes may have increased the likelihood of more extreme variability from the true population mean and therefore may reflect an overestimation of effect.

Limitations

All of the studies included within this review were conducted by the same research team who were also responsible for developing the program. The presence of a vested interest in demonstrating significant positive effects may result in bias in the way that data is analysed and reported. Whilst the published studies demonstrate a level of methodological rigour, it is not possible to ascertain whether any insignificant or negative effects have gone

unreported, something which contributes to the publication bias effect in psychological research (Kühberger, Fritz, & Scherndl, 2014).

Recommendations

Common across the studies, but not the focus of this review, were large effect sizes reported across parental outcome measures. Few studies reported here included longer term follow up of child behavioural outcomes and it is possible that child behaviour change may take longer to become evident after an adapted parenting style has been established. Future research into this program would benefit from not only including a follow up phase but include a longitudinal investigation which could help to provide data on the sustainability of any initial changes observed. Additionally, no research about the effectiveness of the intervention has been undertaken within the UK. Educational Psychology Services may therefore be in an ideal position to pilot the intervention with a view of publishing an evaluation of its efficacy. This would extend the evidence base to within a UK population and could start to counterbalance some of the researcher bias that may be inherent in the published findings thus far.

At the time of writing this review, research was underway evaluating an adapted version for children who had experienced familial trauma. Research evidence suggests that maternal awareness and coaching of negative emotions in their children moderate the relationship between a mother's own symptomology and child behaviour problems in children exposed to domestic violence (Cohodes, Chen, & Lieberman, 2017). The possibility for EPs to

support families exposed to violence through such a program is therefore encouraging. However, researchers and program implementers need to tread carefully and sensitively within this group to ensure that there is no risk of further harm and evidence of the efficacy with this group in particular would need to be sought in a robust manner.

Overall, this review asserts the position that TIK is a promising intervention with the potential to have at least a small yet meaningful impact on children's behaviour. The impact may be more substantial for children with pre-existing behavioural difficulties and these are more likely to be the children referred for EP involvement. The program also has the potential to act as a preventative approach to later behavioural difficulties through having a mediating role in the development of emotional competence. Whilst not the predominant aspect of the EP job role, EPs are in a strong position to work systemically to support parents in developing these knowledge and skills.

5. References

- Bandura, A. (1977). *Social learning theory*. Englewood Cliffs, N.J.: Prentice Hall.
- Barker, C., Pistrang, N., & Elliott, R. (2016). *Research methods in clinical psychology: An introduction for students and practitioners*. (3rd Ed). Chichester: John Wiley & Sons LTD.
- Barnett, A. G., Van der Pols, J. C., & Dobson, A. J. (2005). Regression to the mean: What it is and how to deal with it. *International Journal of Epidemiology*, 34(1), 215–220.
- Bocknek, E., Brophy-Herb, H. E., & Banerjee, M. (2009). Effects of parental supportiveness on toddlers' emotion regulation over the first three years of life in a low-income African American sample. *Infant Mental Health Journal*, 30(5), 452–476.
- Bowlby, J. (1973). *Attachment and loss. Vol. 2: Separation: anxiety and anger*. New York, NY: Basic Books.
- Chen, F. M., Lin, H. S., & Li, C. H. (2012). The Role of Emotion in Parent-Child Relationships: Children's Emotionality, Maternal Meta-Emotion, and Children's Attachment Security. *Journal of Child and Family Studies*, 21(3), 403–410.
- Cohodes, E., Chen, S., & Lieberman, A. (2017). Maternal Meta-Emotion Philosophy Moderates Effect of Maternal Symptomatology on Preschoolers Exposed to Domestic Violence. *Journal of Child and Family Studies*, 26(7), 1831–1843.
- Contreras, J. M., Kerns, K. A., Weimer, B. L., Gentzler, A. L., & Tomich, P. L. (2000). Emotion regulation as a mediator of associations between mother-child attachment and peer relationships in middle childhood. *Journal of Family Psychology*, 14(1), 111–124.
- Denham, S. A., Bassett, H. H., & Wyatt, T. (2007). Socialization of Emotional Competence. *Handbook of Socialization: Theory and Research*, 614–637.
- Duncombe, M. E., Havighurst, S. S., Holland, K. A., & Frankling, E. J. (2012). The contribution of parenting practices and parent emotion factors in children at risk for disruptive behavior disorders. *Child Psychiatry and Human Development*, 43(5), 715–733.
- Dunsmore, J. C., Booker, J. A., & Ollendick, T. H. (2013). Parental emotion coaching and child emotion regulation as protective factors for children with oppositional defiant disorder. *Social Development*, 22(3), 444–466.
- Eisenberg, N., Cumberland, A., & Spinrad, T. L. (1998). Parental Socialization of Emotion. *Psychological Inquiry*, 9(4), 241–273.

- Fan, X. (2001). Statistical significance and effect size in education research: Two sides of a coin. *Journal of Educational Research*, 94(5), 275–282.
- Farrell, P., Woods, K., Lewis, S., Rooney, S., Squires, G., & Connor, M. O. (2006). *A Review of the Functions and Contribution of Educational Psychologists in England and Wales in Light of "Every Child Matters: Change for Children."* Annesley: The Department for Education and Skills Publications.
- Gottman, J. M., & DeClaire, J. (1997). *Raising an emotionally intelligent child*. New York, N.Y.: Simon & Schuster Paperbacks.
- Gottman, J. M., Katz, L. F., & Hooven, C. (1996). Parental meta-emotion philosophy and the emotional life of families: Theoretical models and preliminary data. *Journal of Family Psychology*, 10(3), 243–268.
- 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.
- Gus, L., Rose, J., & Gilbert, L. (2015). Emotion Coaching : A universal strategy emotional and behavioural well-being. *Educational & Child Psychology*, 32(1), 31–41.
- Gutman, L., & Vorhaus, J. (November 2012). The impact of pupil behaviour and wellbeing on educational outcomes childhood wellbeing research centre. *Department for Education: Childgood Wellbeing Research Centre*. Retrieved 25th June 2019, from: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/219638/DFE-RR253.pdf
- Guyatt, G. H., Sackett, D. L., Sinclair, J. C., Hayward, R., Cook, Deborah, J., & Cook, R. J. (1995). Users' Guides to the Medical Literature: IX. A Method for Grading Health Care Recommendations. *The Journal of the American Medical Association*, 274(22), 1800–1804.
- Havighurst, S.S., Duncombe, M., Frankling, E., Holland, K., Kehoe, C., & Stargatt, R. (2015). An Emotion-Focused Early Intervention for Children with Emerging Conduct Problems. *Journal of Abnormal Child Psychology*, 43, 749–760
- Havighurst, S. S., Kehoe, C. E., & Harley, A. E. (2015). Tuning in to Teens: Improving Parental Responses to Anger and Reducing Youth Externalizing Behavior Problems. *Journal of Adolescence*, 42, 148–158
- Havighurst, S. S., Kehoe, C. E., Harley, A. E., & Wilson, K. R. (2015). Tuning in to Kids : an emotion-focused parenting intervention for children with disruptive behaviour problems. *Making Parenting Work for Children's Mental Health*, 41–50.
- Havighurst, S. S., Wilson, K. R., Harley, A. E., & Prior, M. R. (2009). Tuning in to Kids: An emotion focused parenting program - initial findings from a

- community trial. *Journal of Community Psychology*, 37(8), 1008–1023.
- Havighurst, S., Wilson, K., Harley, A., Kehoe, C., Efron, D., & Prior, M. (2013). "Tuning into Kids": Reducing Young Children's behavior Problems Using an Emotion Coaching Parenting Program. *Child Psychiatry and Human Development*, 44(2), 247-264
- Johnson, A. M., Hawes, D. J., Eisenberg, N., Kohlhoff, J., & Dudeney, J. (2017). Emotion socialization and child conduct problems: A comprehensive review and meta-analysis. *Clinical Psychology Review*, 54, 65–80.
- Katz, L. F., Maliken, A. C., & Stettler, N. M. (2012). Parental Meta-Emotion Philosophy: A Review of Research and Theoretical Framework. *Child Development Perspectives*, 6(4), 417–422.
- Kratochwill, T. R. (2003). Evidence-Based Practice: Promoting Evidence-Based Interventions in School Psychology. *School Psychology Quarterly*, 18(4), 389-408.
- Kühberger, A., Fritz, A., & Scherndl, T. (2014). Publication bias in psychology: A diagnosis based on the correlation between effect size and sample size. *PLoS ONE*, 9(9), 1–8.
- Linden, A. (2013). Assessing regression to the mean effects in health care initiatives. *BMC Medical Research Methodology*, 13(1), 1-7
- Loop, L., & Roskam, I. (2016). Do Children Behave Better When Parents' Emotion Coaching Practices are Stimulated? A Micro-Trial Study. *Journal of Child and Family Studies*, 25(7), 2223–2235.
- Mcguiggan, C. L. (2017). Stepping over the boundary : An exploration of educational psychologists' work with families. *Doctoral Thesis*
- Meybodi, F. A., Mohammadkhani, P., Pourshahbaz, A., Dolatshahi, B., & Havighurst, S. (2017). Research paper: Reducing children behavior problems: A pilot study of tuning in to kids in Iran. *Iranian Rehabilitation Journal*, 15(3), 269–276
- Morris, A. S., Silk, J. S., Steinberg, L., Myers, S. S., & Robinson, L. R. (2007). The role of the family context in the development of emotion regulation. *Social Development*, 16(2), 361–388.
- Noble, T., & McGrath, H. (2012). Wellbeing and resilience in young people and the role of positive relationships. In S. Roffey (Ed.), *Positive Relationships: Evidence based practice across the world* (pp. 17–33). New York, NY: Springer.
- Osei, D. (2015). *Does the 'Tuning into Kids' program enable parents to reduce emotional and behavioural difficulties in their children? An evidence-based practice report*. Submitted as part of the Doctorate in Educational and Child Psychology. London; University College London.

- Sheeber, L., Shortt, J. W., Low, S., & Katz, L. F. (October, 2010). *Emotion coaching as a unique predictor of adolescent internalizing problems*. Paper presented at the annual meeting of the Society for Research in Psychopathology, Seattle, WA.
- Shortt, J. W., Stoolmiller, M., Smith-Shine, J. N., Mark Eddy, J., & Sheeber, L. (2010). Maternal emotion coaching, adolescent anger regulation, and siblings' externalizing symptoms. *Journal of Child Psychology and Psychiatry*, 51(7), 799–808.
- 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.
- Stringer, P., Powell, J., & Burton, S. (2006). Developing a community psychology orientation in an educational psychology service. *Educational and Child Psychology, Community Psychology: Theoretical and Practical Implications for Educational Psychologists*, 23(1), 59-67
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press
- Wilson, K., Havighurst, S. S., & Harley, A. E. (2014). Dads Tuning in to Kids: Piloting a new parenting program targeting fathers' emotion coaching skills. *Journal of Community Psychology*, 42(2), 162–168
- Wilson, K. T., Havighurst, S., S. Kehoe, C. E., and Harley, A. E. (2016). Dads Tuning in to Kids: Preliminary Evaluation of a Father's Parenting Program. *Family Relations*, 65, 535–549
- Wong, M. S., McElwain, N. L., & Halberstadt, A. G. (2009). Parent, Family, and Child Characteristics: Associations With Mother- and Father-Reported Emotion Socialization Practices. *Journal of Family Psychology*, 23(4), 452–463.

6. Appendices

Appendix 1. Excluded studies and the reasons for exclusion

Studies	Reason for exclusion
Duncombe, M. E., Havighurst, S. S., Holland, K. A., Frankling, E. J., Kehoe, C., & Stargatt, R. (2014). Comparing an emotion- and behavior-focused group parenting program as part of multisystemic intervention for child conduct disorder. <i>Journal of Child and Adolescent Psychology, 45</i> (3), 320–334	Uses data from a different study (2a)
Dunsmore, J. C., Booker, J. A., Ollendick, T. H., & Greene, R. W. (2016). Emotion Socialization in the Context of Risk and Psychopathology: Maternal Emotion Coaching Predicts Better Treatment Outcomes for Emotionally Labile Children with Oppositional Defiant Disorder. <i>Social Development, 25</i> (1), 8–2	Not the 'Tuning into Kids' program (3b) Mothers assessed for emotion coaching but this was not targeted through the interventions chosen.
Havighurst, S. S., Harley, A., & Prior, M. (2004). Building preschool children's emotional competence: A parenting program. <i>Early Education and Development, 15</i> (4), 423–448	Undertaken prior to 2013 (7) Included in a previous review
Havighurst, S. S., Wilson, K. R., Harley, A. E., & Prior, M. R. (2009). Tuning in to kids: An emotion-focused parenting program – initial findings from a community trial. <i>Journal of Community Psychology, 37</i> (8), 1008–1023.	Undertaken prior to 2013 (7) Included in a previous review
Havighurst, S. S., Wilson, K. R., Harley, A. E., Prior, M. R., & Kehoe, C. (2010). Tuning in to Kids: Improving emotion socialization practices in parents of preschool children – findings from a community trial. <i>Journal of Child Psychology and Psychiatry, 51</i> (12), 1342–1350.	Undertaken prior to 2013 (7) Included in a previous review
Herbert, S. D., Harvey, E. A., Roberts, J. L., Wichowski, K., & Lugo-Candelas, C. I. (2013). A Randomized Controlled Trial of a Parent Training and Emotion Socialization Program for Families of Hyperactive Preschool-Aged Children. <i>Behavior Therapy, 44</i> (2), 302–316	Not the 'Tuning into Kids' program (3b)
Isobel, S., Meehan, F., & Pretty, D. (2016). An Emotional Awareness Based Parenting Group	Reported on parental

Studies	Reason for exclusion
for Parents with Mental Illness: A Mixed Methods Feasibility Study of Community Mental Health Nurse Facilitation. <i>Archives of Psychiatric Nursing</i> , 30(1), 35–40	outcomes only (4c)
Kehoe, C. E. (2014). Tuning in to Teens: Examining the efficacy of an emotion focused parenting intervention in reducing pre-adolescents' internalising difficulties. <i>Doctoral Thesis</i> .	Uses data from a different study (2a) Outcomes focus on internalizing behaviours (4b) An unpublished doctoral thesis (1)
Kehoe, C. E., Havighurst, S. S., & Harley, A. E. (2014). Tuning in to Teens: Improving parent emotion socialization to reduce youth internalizing difficulties. <i>Social Development</i> , 23(2), 413–431.	Outcomes focus on internalizing behaviours (4b)
Lauw, M. S. M., Havighurst, S. S., Wilson, K., Harley, A. E., & Northam, E. A. (2014). Improving parenting of toddlers' emotions using an emotion coaching parenting program: A pilot study of tuning in to toddlers. <i>Journal of Community Psychology</i> , 42(2), 169–175.	Toddler age group (5b) Included in a previous review
Loop, L., Mouton, B., Stievenart, M., & Roskam, I. (2017). One or many? Which and how many parenting variables should be targeted in interventions to reduce children's externalizing behavior? <i>Behaviour Research and Therapy</i> , 92, 11–23	Not the 'Tuning into Kids' program (3b)
Webster-Stratton, C. H., Jamila Reid, M., & Beauchaine, T. (2011). Combining parent and child training for young children with ADHD. <i>Journal of Clinical Child and Adolescent Psychology</i> , 40(2), 191–203	Not the 'Tuning into Kids' program (3b)
Wilson, K. R., Havighurst, S. S., & Harley, A. E. (2012). Tuning in to Kids: An effectiveness trial of a parenting program targeting emotion socialization of preschoolers. <i>Journal of Family Psychology</i> , 26(1), 56–65.	Undertaken prior to 2013 (7) Included in a previous review

Appendix 2. List of included studies for the systematic literature review

	Studies
1	Havighurst, S.S., Duncombe, M., Frankling, E., Holland, K., Kehoe, C., & Stargatt, R. (2015). An Emotion-Focused Early Intervention for Children with Emerging Conduct Problems. <i>Journal of Abnormal Child Psychology</i> , 43, 749–760
2	Havighurst, S. S., Kehoe, C. E., & Harley, A. E. (2015). Tuning in to Teens: Improving Parental Responses to Anger and Reducing Youth Externalizing Behavior Problems. <i>Journal of Adolescence</i> , 42, 148–158
3	Havighurst, S., Wilson, K., Harley, A., Kehoe, C., Efron, D., & Prior, M. (2013). "Tuning into Kids": Reducing Young Children's behavior Problems Using an Emotion Coaching Parenting Program. <i>Child Psychiatry and Human Development</i> , 44(2), 247-264
4	Meybodi, F. A., Mohammadkhani, P., Pourshahbaz, A., Dolatshahi, B., & Havighurst, S. (2017). Research paper: Reducing children behavior problems: A pilot study of tuning in to kids in Iran. <i>Iranian Rehabilitation Journal</i> , 15(3), 269–276.
5	Wilson, K., Havighurst, S. S., & Harley, A. E. (2014). Dads Tuning in to Kids: Piloting a new parenting program targeting fathers' emotion coaching skills. <i>Journal of Community Psychology</i> , 42(2), 162–168
6	Wilson, K. T., Havighurst, S., S. Kehoe, C. E., and Harley, A. E. (2016). Dads Tuning in to Kids: Preliminary Evaluation of a Father's Parenting Program. <i>Family Relations</i> , 65, 535–549

Appendix 3. Mapping the Field

Authors	Study design	Study location	Sample characteristics	Intervention details	Key outcome measures ¹	Key findings ²
1. Havighurst and Duncombe et al. (2015)	Random assignment to intervention or control condition at the school level. Outcome measures completed at baseline and at 10 months post (6 months?).	Community Participants selected from schools in lower SES, metropolitan regional and rural areas. Victoria, Australia	Sample size at baseline: 231 primary caregivers, intervention = 113 and control = 118 Target children: aged between 5 and 9 with emerging conduct problems as screened by a score of 1 or greater on the Conduct Problems Risk Screen. Female = 26% , male = 74% Parents: female = 96%, male = 4%. Mean age: 36.8 years. Mean working hours: 14.3 hours per week. Marital status: 63.5% married or in marital type relationship (defacto). Percentage of non-high school completion: 48.6%. Household income: 60% below the mean household income for the state (\$63, 232).	Tuning in to Kids (TIK) Eight 2 hour weekly group parenting sessions	Parents: The Eyberg Child Behaviour Inventory 6 (ECBI) Teacher: The Strengths and Difficulties Questionnaire (SDQ) and the Social Competence Rating Scale	Parents reported statistically significant reductions in behaviour problems within the intervention group as opposed to the control group. Teachers reported statistically significant reductions in behaviour problems within the intervention group as opposed to the control group.

Authors	Study design	Study location	Sample characteristics	Intervention details	Key outcome measures ¹	Key findings ²
2. Havighurst, Kehoe and Harley, (2015)	Random assignment to intervention or control condition at the school level. Outcome measures completed at baseline and at 10.5 months post baseline	Community Metropolitan Melbourne, Australia	Sample size at baseline: 225 primary caregivers, intervention = 121 and control = 104 Target children: aged between 10 and 13 years. Female = 51%, male = 49% Parents: female = 89%, male = 11%. Mean age: 44.1 years. Mean working hours: 29.5 hours per week. Percentage marital status: 82.7% married or defacto. Percentage of non-high school completion: 24.4%. Household income: 38.7% combined income of \$100,000 or greater.	Tuning in to Teens (TINT) Six 2 hour weekly group parenting sessions	The Conduct Problems and the Hyperactivity-Inattention subscales of the Strengths and Difficulties Questionnaire completed by parents and the young people.	Significant reductions were observed in youth externalising difficulties for the intervention group as rated by both parents and young people with small effect sizes noted
3. Havighurst et al. (2013)	Random assignment to intervention or comparison condition (treatment as	Clinical Behaviour Clinic of the Royal Children's Hospital and the Western Sunshine	Sample size at baseline: 54 primary caregivers, intervention = 31 and control = 23 Target children: aged between 4 and 5 years with elevated scores on the ECBI (above the clinical cut off). Female = 22%,	TIK Six 2 hour weekly group parenting sessions and two booster	Parents: The Eyberg Child Behavior Inventory Teachers: The Sutter-Eyberg Student Behaviour	Parents in the intervention condition reported significant reductions in child behaviour intensity.

Authors	Study design	Study location	Sample characteristics	Intervention details	Key outcome measures ¹	Key findings ²
	normal). Outcome measures completed at baseline, post and 6 month follow up.	Hospital Melbourne Australia NB. Parents of preschool children (non-clinical) were also included within the groups but were not part of the research.	male = 78% Parents: female = 100%, male = 0%. Mean age: 35.66 years. Mean working hours: not reported. Marital status: not reported. Percentage of non-high school completion: 75.5%. Household income: 46% combined income of \$40,000 – 99,999	sessions at two monthly intervals thereafter.	Inventory (S-ESBI)	Teachers of children in the intervention condition reported reductions in behaviour intensity and fewer problems with child behaviour than those in the waitlist control condition.
4. Meybodi et al. (2017)	Random assignment to intervention or control condition at the school level. Outcome measures completed at	Community Diverse lower to upper-class socio-economic areas of Tehran Iran	Sample size at baseline: 54 primary caregivers, intervention = 27 and control = 27 Target children: aged between 3 and 6 years with behaviour problems as identified by a T score of 65 or higher on the externalizing subscale of the Child Behaviour Check-list (CBCL). Parents: female = 100%, male =	TIK Six 2 hour weekly group parenting sessions and two booster sessions at two monthly intervals	The Eyberg Child Behaviour Inventory	Mothers within the intervention group reported significant reductions in child behaviour problems with a large effect size that continued at follow up as noted by the ECBI problem score.

Authors	Study design	Study location	Sample characteristics	Intervention details	Key outcome measures ¹	Key findings ²
	baseline, post and 3 months follow up.		0%. Mean age: 34.21 years. Mean working hours: not reported. Marital status: 100% married. Percentage of non-high school completion: 2.1%. Household income: family income ranged from 1 million to over 5 million Tomans.	thereafter.		
5. Wilson, Havighurst and Harley (2014)	Single group pretest-posttest design. Outcome measures completed at baseline and post.	Community Preschools in the state of Victoria, Australia	Sample size at baseline: 43 Target children: aged between 3 and 5 years. Parents: female = 0%, male = 100%. Mean age: 41.05 years. Mean working hours: not reported. Marital status: not reported. Percentage of non-high school completion: 9.3%. Household income: 62.8% combined income of \$100,000 or greater.	Dads Tuning in to Kids (DadsTIK) Seven 2 hour weekly group parenting sessions.	The Strengths and Difficulties Questionnaire	Fathers reported reductions in difficult child behaviours at post-test as noted by the SDQ: Total difficulties score.
6. Wilson et al. (2016)	Random assignment to intervention	Preschools within a 15-km radius of the University of	Sample size at baseline: 162 intervention = 87 and control = 75	DadsTIK Seven 2 hour weekly	The Strengths and Difficulties Questionnaire	Fathers within the intervention group reported small but statistically

Authors	Study design	Study location	Sample characteristics	Intervention details	Key outcome measures ¹	Key findings ²
	or control condition at the school level. Outcome measures completed at baseline and post.	Melbourne and six preschools in Geelong (Victoria) Australia	Target children: aged between 3 and 6 years. Female = 46.3%, male = 53.7% Parents: female = 0%, male = 100%. Mean age: 40.6 years. Mean working hours: 41.8 hours per week. Percentage marital status: 97.5% married or defacto. Percentage of non-high school completion: 9.3%. Household income: 76.5% combined income of \$100,000 or greater.	group parenting sessions with a booster session offered 6 to 8 weeks later		significant reductions in difficult child behaviours as noted by the SDQ: Total difficulties score.

¹ & ² Key outcomes and findings reported here relate to the specific purpose of the review question and as such, only measures and findings that relate to child behavioural outcomes are reported within this review.

Appendix 4. Weight of Evidence A

Appendix 4a. Modifications to the Kratochwill (2003) coding protocol

APA Task Force Coding Protocol by Kratochwill (2003) coding protocol was used to code each of the studies in order to generate a 'Weight of Evidence A' rating for included studies. The table below shows the adaptations made to the protocol along with a rationale for these amendments.

Items removed	Rationale
Part I	
Sections B.7 - B.8. Coding for qualitative research methods	Studies did not use qualitative research methods.
Part II	
Section C. Primary/Secondary Outcomes Are Statistically Significant	The quality of the methodology is evaluated using this protocol and outcomes are considered separately within the review.
Section D. Educational/Clinical Significance	The majority of the studies either examined the program as a universal or preventative intervention or when used as an intervention for behavioural difficulties, these were either not at clinically diagnosis levels or clinical diagnosis levels were not reported. Only one study (Havighurst et al., 2013) used a clinical sample and reported on this aspect.
Section E	The intervention is manualised and components are not separated.
Section G	There was no within study replication.
Section H. Site of Implementation	Site of implementation was not deemed relevant to the review question nor would it add significant value in terms of reviewing the quality of the methodology of the studies.

Items removed	Rationale
Part III	
Section A4. Receptivity/ acceptance by target population	Receptivity of intervention by target group was not deemed necessary in determining methodological quality.
Section B. Length of Intervention	Specified elsewhere within the review.
Section C. Intensity/Dosage of Intervention	Specified elsewhere within the review.
Section D. Dosage Response	The intervention is a manualised program with compulsory content over a set number of sessions. It was therefore not possible for effects of a higher dosage to be observed.
Section F. Characteristics of the Intervener	None of the studies reported this detail.
Section G. Intervention style of Orientation	All studies focus on the same intervention as per inclusion criteria. Theoretical basis of intervention is outline separately in the review.
Section H. Cost Analysis Data	Not reported and unnecessary for the purpose of reviewing methodological quality.
Section I. Training and Support Resources	Not reported and unnecessary for the purpose of reviewing methodological quality.
Section J. Feasibility	Not reported and unnecessary for the purpose of reviewing methodological quality.

Appendix 4b. Sample coding protocol

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

Coding Protocol

Name of Coder: Hayley Prickett

Date: 10.01.19

Full Study Reference in proper format:

Meybodi, F. A., Mohammadkhani, P., Pourshahbaz, A., Dolatshahi, B., Havighurst, S. (2017). Reducing Children Behavior Problems: A Pilot Study of Tuning in to Kids in Iran. *Iranian Rehabilitation Journal*. 15(3), 269 - 275

Intervention name: (description of study): Tuning in to Kids

Type of Publication:

- Book/Monograph
- Journal Article
- Book Chapter
- Other (specify):

I. 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 classes)
- 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

- Current exposure
- Prior exposure
- Unknown

II. 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 (answer A1 through A4)

A1. Use of outcome measures that produce reliable scores for the majority of primary outcomes (select one of the following)

- Yes
- No
- Unknown/unable to code

High Chronbach's alphas reported for all measures (0.76 – 0.92 with 0.92 for the outcome measure of interest.

A2 Multi-method (select one of the following)

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

Three measures administered to parent with one being related to child outcomes

A3 Multi-source (select one of the following.)

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

All parental reports

A4 Validity of measures reported (select one of the following)

- Yes validated with specific target group (ECBI and Parent Emotional Style Questionnaire)
- In part, validated for general population only (General Health Questionnaire)

- No
- Unknown/unable to code

Rating for measurement (select 0, 1, 2 or 3) 3 2 1 0

Variety of assessment measures but only assessed by one source (not triangulated). Generally acceptable or high levels of internal consistency reliability of the measures at both time points as reported through Chronbach's alpha scores. Some reported details about the validity of the scales either generally or with the target population.

B. Comparison Group

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

- Typical contact
- 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

- By change agent
- Statistical
- 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

T-tests and chi square analysis was used to examine baseline differences, there were no significant differences between groups on any of the demographic or outcome variables suggesting that the randomisation had resulted in comparable groups.

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 (select 0, 1, 2 or 3)

3 2 1 0

Presence of a control group (but not an active treatment), randomised allocation between groups at the individual level, post hoc test group equivalence analysis.

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

Supervision of delivery by lead program developer and fidelity checklists were completed after every session to ensure the program was delivered according to the manual.

F2. Manualization (select all that apply)

- Written material involving a detailed account of the exact procedure and the sequence 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.

Both facilitators trained by the TIK first author (Havighurst).

Rating for Implementation Fidelity (select 0, 1, 2 or 3):

3 2 1 0

I. Follow Up Assessment

Timing of follow up assessment: 3 months only (only one time point for follow up)

Number of participants included in the follow up assessment: specify __48? _____

Consistency of assessment method used: specify: **Same measures administered**

Rating for Follow Up Assessment (select 0, 1, 2, or 3): 3 2 1 0

III. Other Descriptive or Supplemental Criteria to Consider

A. External Validity Indicators

A1. Sampling procedures described in detail Yes No

Specify rationale for selection: **Inclusion and exclusion criteria set based on the objectives of the study, i.e. parents of pre-school age children with behaviour problems.**

Specify rationale for sample size: **Initial sample size was larger but some were screened out because they did not meet the inclusion criteria. All that consented to participate and met the criteria were included.**

A1.1 Inclusion/exclusion criteria specified? Yes No

A1.2 Inclusion/exclusion criteria similar to school practice
Yes No

A1.3 Specified criteria related to concern
Yes No

A3. Details are provided regarding variables that:

A3.1 Have differential relevance for intended outcomes

Yes No

Specify: NA?

A3.2 Have relevance to inclusion criteria Yes No

Specify: **mothers of pre-school attending children, children with behavioural concerns as indicated by a T score of 45 or higher on the externalizing subscale of the CBCL, children without an intellectual or pervasive developmental disorder**

A5. Generalization of Effects:

A5.1 Generalization over time

A5.1.1 Evidence is provided regarding the sustainability of outcomes after intervention is terminated

Yes No

Specify: Three month follow up with effect sizes reported to suggest a continued effect of treatment after that time period for the ECBI only (measure of interest). Extending beyond that to after the child started school was suggested as a way of providing further evidence as to the efficacy of the treatment.

A5.1.2 Procedures for maintaining outcomes are specified

Yes No

Specify: _____

A5.2 Generalization across settings

A5.2.1 Evidence is provided regarding the extent to which outcomes are manifested in contexts that are different from the intervention context

Yes No

Specify: _____

A5.2.2 Documentation of efforts to ensure application of intervention to other settings

Yes No

Specify: _____

A5.2.3 Impact on implementers or context is sustained

Yes No

Specify: _____

A5.3 Generalization across persons

Evidence is provided regarding the degree to which outcomes are manifested with participants who are different than the original group of participants for with the intervention was evaluated Yes No

Specify: _____

E. Program Implementer (select all that apply)

- E1. Research Staff
- E2. School Specialty Staff
- E3. Teachers
- E4. Educational Assistants
- E5. Parents
- E6. College Students
- E7. Peers
- E8. Other
- E9. Unknown/insufficient information provided.

Summary of Evidence

Indicator	Overall evidence rating 0-3	Description of evidence Strong Promising Weak No/limited evidence Or Descriptive ratings
General Characteristics		
Design	NNR	
Statistical treatment/data analysis	NNR	
Type of programme	NNR	
Stage of programme	NNR	
Concurrent/ historical intervention exposure	NNR	
Key features		
Measurement	2	Promising
Comparison group	2	Promising
Implementation fidelity	3	Strong
Follow up	2	Promising

Appendix 4c. Coding criteria

Through the Kratochwill (2003) coding protocol, studies were weighted on four key dimensions: 'measurement', 'comparison group', 'implementation fidelity', and 'follow-up assessment' with each dimension receiving a corresponding numerical value. Tables 1a – 1d provide the criteria for judging each dimension as either high (strong evidence), medium (promising evidence), low (weak evidence) or none/not reported (no evidence) and Table 2 details the score attributed to each study on each dimension.

Table 1a.

Measures

Weighting	Descriptors
High (3)	<ul style="list-style-type: none">• Study used measures that produce reliable scores of at least .85, for the majority of primary outcomes• Study provides details about the type of reliability statistics used• Data was collected using multiple methods and collected from multiple sources• Validity is reported.
Medium (2)	<ul style="list-style-type: none">• Study used measures that produce reliable scores of at least .70 for the primary outcomes• Data was collected using multiple methods and/or collected from multiple sources• A case for validity does not need to be presented.
Low (1)	<ul style="list-style-type: none">• Study used measures that produce reliable scores of at least .50 for the primary outcomes• Data may have been collected using multiple methods and/or multiple sources although this is not a necessity.• A case for validity does not need to be presented.
None or not reported (0)	<ul style="list-style-type: none">• None of the above criteria is met OR there is no explicit reported evidence in relation to the criteria.

Table 1b.

Comparison group

Weighting	Descriptors
High (3)	<ul style="list-style-type: none"> • Uses at least one type of "active" comparison group • Initial group equivalency must be established OR • Where group equivalency is not established, appropriate analyses are conducted to account for any differences found (e.g. an analysis of covariance). • Evidence that change agents were counterbalanced (where relevant) and that effect of class/school has been controlled for. • Equivalent mortality and low attrition at post, and if applicable, at follow-up.
Medium (2)	<ul style="list-style-type: none"> • Uses at least a "no intervention group" type of comparison • There is evidence for at least two of the following: counterbalancing of change agents, group equivalence established or accounted for, or equivalent mortality with low attrition. • If equivalent mortality is not demonstrated, an intent-to intervene analysis is conducted which demonstrated no significant group differences between the control and the treatment group.
Low (1)	<ul style="list-style-type: none"> • The study uses a comparison group • No group equivalence procedures were used • At least one of the following is present: counterbalancing of change agents, group equivalence established, or, equivalent mortality with low attrition. • If equivalent mortality is not demonstrated, an intent-to intervene analysis is conducted.
None or not reported (0)	<ul style="list-style-type: none"> • None of the above criteria is met OR there is no explicit reported evidence in relation to the criteria.

Table 1c.

Fidelity

Weighting	Descriptors
High (3)	<ul style="list-style-type: none"> • The study demonstrates strong evidence of acceptable adherence. • Evidence of fidelity is measured through at least two of the following: ongoing supervision/consultation, coding sessions or audio/video tapes, and a use of a manual. • The “manual” is either written materials involving a detailed account of the exact procedures and the sequence in which they are to be used or formal training session detailing exact procedures and sequence.
Medium (2)	<ul style="list-style-type: none"> • The study demonstrates evidence of acceptable adherence. • Evidence of fidelity is measured through at least one of the following: ongoing supervision/consultation, coding sessions, or audio/video tapes, and use of a manual. • The “manual” is either written materials involving an overview of broad principles and a description of the intervention phases, or formal/informal training session involving an overview of broad principles and a description of the intervention phases.
Low (1)	<ul style="list-style-type: none"> • Demonstrates evidence of acceptable adherence measured through at least one of the above criteria or use of a manual.
None or not reported (0)	<ul style="list-style-type: none"> • None of the above criteria is met OR there is no explicit reported evidence in relation to the criteria.

Table 1d.

Follow-up

Weighting	Descriptors
High (3)	<ul style="list-style-type: none"> • The study conducted follow up assessments over multiple intervals with all participants that were included in the original sample. • Similar measures used to analyse data from primary or secondary outcomes.
Medium (2)	<ul style="list-style-type: none"> • The study conducted follow up assessments at least once with the majority of participants that were included in the original sample. • Similar measures used to analyse data from primary or secondary outcomes.
Low (1)	<ul style="list-style-type: none"> • The study conducted follow up assessments at least once with some participants from the original sample.
None or not reported (0)	<ul style="list-style-type: none"> • None of the above criteria is met OR there is no explicit reported evidence in relation to the criteria.

Table 2.

WoE A study comparison

Studies	Measures	Comparison group	Fidelity	Follow up	Overall quality of methodology*
1. Havighurst et al. (2015)	3	2	3	0	Medium (2.00)
2. Havighurst et al. (2015)	2	2	3	0	Medium (1.75)
3. Havighurst et al.	3	2	3	2	High (2.50)

Studies	Measures	Comparison group	Fidelity	Follow up	Overall quality of methodology*
(2013)					
4. Meybodi et al. (2017)	2	2	3	2	Medium (2.25)
5. Wilson, Havighurst and Harley (2014)	2	NNR	1	0	Low (1.00)
6. Wilson et al. (2016)	2	2	3	0	Medium (1.75)

*Low = <1.4, Medium = 1.5 – 2.4, High = >2.5

Appendix 5. Weight of Evidence B

This weighting is a review-specific judgement about the suitability of the evidence for answering the review question (Gough, 2007). In this case, whether the research design is appropriate for evaluating the effectiveness of the Tuning in to Kids program in reducing child behaviour difficulties (externalizing). Criteria for WoE B decisions are guided by evidence hierarchies that identify Randomised Controlled Trials (RCTs) as the most robust methodology for evaluating the effectiveness of an intervention (Guyatt et al., 1995).

The rationale for the criteria used in this weighting is expanded on below:

1. *Control group*: a control enables the analysis of intervention effect by comparing any changes in the intervention group with a group who have not had access to the intervention. An active control group

enhances this as it compares the effect of the intervention with an alternative treatment or treatment as normal, whereas a wait list control group can only demonstrate that receiving the intervention is better than no treatment at all.

2. *Randomisation of allocation*: randomly allocating participants to either the treatment or control group ensures that any between participant differences are minimised and thus reduces the possibility for bias in the sample/reduces the influence of any confounding variables. This is stronger when participants are randomly allocated on an individual rather than at a group (e.g. school) level. Randomisation at the group level applies statistical tests of group equivalence with any group differences being included in post-hoc analyses.
3. *Pre and post intervention measures*: data collection before and after the intervention allows for an analysis of the changes within the participants in relation to the intervention. Including a longer term follow up measure allows for an assessment of the longevity of intervention effects over time. For this review question it is beneficial to consider whether the impact on a child's behaviour is sustained after the intervention has ceased.
4. *Sufficient sample size*: this enables the analysis of the intervention effect to be separated from the effect of other variables and to thus be able to confirm whether the results are statistically significant, i.e. unlikely to have occurred by chance.

Table 1.

Criteria for WoE B

All but one of the criterion within each range must be met.

Weighting	Criteria
High (3)	<ul style="list-style-type: none"> • The study must have an 'active' control group. • Participants must be randomly allocated to condition at the individual level. • Group equivalence analysed and post hoc tests adjust for the effects of any differences identified. • The study must collect pre and post measures for relevant outcomes (child outcomes). • The study must have a sample size that is adequate for all statistical analysis*.
Medium (2)	<ul style="list-style-type: none"> • The study must have at least a 'no intervention' control group • Participants are randomly allocated to treatment or control although randomisation may be done at the group level (e.g. school or preschool). • Group equivalence must have been established post hoc (with appropriate analyses and adjustments made). • Where multiple interventions occur, dismantling procedures are used to identify the effects of the intervention in question specifically. • The study must collect pre and post measures for relevant outcomes (child outcomes). • The study may have a sample size that is no more than 25% smaller than required for statistical analysis.
Low (1)	<ul style="list-style-type: none"> • The study does not include a control group. • The study must collect pre and post measures for relevant outcomes. • The study may have a sample size that is lower than required for statistical analysis. • Where multiple interventions occur, no dismantling procedures are used in which to identify the effects of the intervention in question specifically.

Zero (0)	<ul style="list-style-type: none"> The study does not meet any of the aforementioned criteria
----------	--

*Adequate sample size was determined from Cohen (1992) based on a medium effect size (0.5), alpha level of 0.05 and a power of 0.8.

Table 2.

Scores for WoE B

Studies	Overall WoE B
1. Havighurst and Duncombe et al. (2015)	Medium (2)
2. Havighurst, Kehoe, and Harley(2015)	Medium (2)
3. Havighurst et al. (2013)	Medium (2)
4. Meybodi et al. (2017)	Medium (2)
5. Wilson, Havighurst and Harley (2014)	Low (1)
6. Wilson et al. (2016)	Medium (2)

Appendix 6. Weight of Evidence C

This criteria for WoE C was determined based on relevance to the review question, i.e. the extent to which the study could demonstrate effectiveness of the TIK program for children and the relevance and thus generalisability of this within EP practice in the UK.

Table 1.

Criteria for WoE C

Weighting	Descriptors
High (3)	<p>At least five out of the six criteria below are met:</p> <ol style="list-style-type: none"> Study takes place within the UK or in a country where the education system is likely to be comparable and the language of instruction is English (e.g. USA,

	Canada, Australia, New Zealand).
	2. Study selects a sample that represents a diversity of socio economic groups amongst parents with less than 30% in the highest reported household income bracket.
	3. Study selects the primary caregiver, regardless of whether this is male or female.
	4. Children are drawn from non-clinical settings, e.g. preschools and schools rather than behaviour clinics/hospital settings.
	5. Study selects pupils that are 'at risk' for conduct difficulties based on a behaviour measure screening process.
	6. Primary outcome measures focus on outcomes for children OR where outcome measures for children are secondary outcomes, at least two methods (e.g. questionnaire, behaviour rating scale, observation) or at least two sources (e.g. parent/teacher/child) are used to measure child outcomes.
Medium (2)	At least three of the above criteria are met
Low (1)	At least one of the above criteria are met
Zero (0)	The study meets either one OR none of the criteria outlined above.

Table 2.

Scores for WoE C

Studies	Overall WoE C	Criteria met
1. Havighurst et al. (2015)	High (3)	1, 2, 3, 4, 5 and 6
2. Havighurst et al. (2015)	Medium (2)	1, 3, 4 and 6
3. Havighurst et al. (2013)	High (3)	1, 2, 3, 5 and 6
4. Meybodi et al. (2017)	Medium (2)	3, 4 and 5
5. Wilson, Havighurst and Harley (2014)	Low (1)	1 and 4
6. Wilson et al. (2016)	Low (1)	1 and 4
