The Parent Programme Implementation Checklist (PPIC): The development and testing of an objective measure of skills and fidelity for the delivery of parent programmes

Prepared for submission to Child Care in Practice

Tracey Bywater[1], Nicole Gridley[2], Vashti Berry[3], Sarah Blower[2] Kate Tobin[4]

Abstract Word Count: 299/300
Word Count: 7165/up to 7500 (exclude references and tables)
Keywords: Parent Programmes/ Intervention/ Implementation/ Fidelity/ Reliability/ Observation/ Measurement

[1] Corresponding Author: Department of Health Sciences, University of York, YO10 5DD. Tel 01904 328105
[2] Department of Health Sciences, University of York, YO10 5DD
[3] University of Exeter Medical School, Exeter, EX1 2LU
[4] Dartington Social Research Unit, Glasgow, G2 4TB

Abstract

Background: Group-based parent programmes demonstrate positive benefits for adult and child mental health, and child behaviour outcomes. Greater fidelity to the programme delivery model equates to better outcomes for families attending, however, fidelity is typically self-monitored using programme specific checklists. Self-completed measures are open to bias, and it is difficult to know if positive outcomes found from research studies will be maintained when delivered in regular services. Currently, ongoing objective monitoring of quality is not conducted during usual service delivery. This is odd given that quality of other services is assessed objectively, e.g. OFSTED. Independent observations of programme delivery are needed to assess fidelity and quality of delivery to ensure positive outcomes, and therefore justify the expense of programme delivery.

Methods: This paper outlines the initial development and reliability of a tool, The Parent Programme Implementation Checklist (PPIC) which was originally developed as a simple, brief and generic observational tool for independent assessment of implementation fidelity of group-based parent programmes. PPIC does not require intensive observer training before application/use. This paper presents initial data obtained during delivery of the Incredible Years BASIC programme across 9 localities in England and Wales.

Results: Reasonable levels of inter-rater reliability were achieved across each of the three subscales (Adherence, Quality and Participant Responsiveness) and the overall total score when applying percentage agreements (>70%) and intra-class correlations (*ICC* range between 0.404 and 0.730). Intra-rater reliability ($n = 6$) was acceptable at the subscale level.

Conclusions: We conclude that the PPIC has promise, and with further development could be utilised to assess fidelity of parent group delivery during research trials and standard service delivery. Further development would need to include data from other parent programmes, and testing by non-research staff. The objective assessment of quality of delivery would inform services where improvements could be made.

Introduction

Systematic reviews and meta-analyses demonstrate that cognitive-behavioural group-based parenting programmes are effective in improving parental mental health, parenting skills, and child pro-social behaviour for parents and their children, aged three to twelve years, who are at risk of developing conduct disorder (e.g. Barlow et al., 2010; 2012; 2014; Furlong et al. 2012). Despite this, it is not always clear if it is the intervention itself, the process of programme implementation, or the combination and interaction of both elements that influences these outcomes (Axford et al., 2017; Bywater, 2012). The Medical Research Council (MRC, Moore et al., 2008; Moore et al., 2015) stipulates that complex interventions such as parent programmes, comprise several interacting components that impact on familial outcomes. During initial implementation a good theoretical understanding of the programme is needed to assess the impact on behaviour change, so that weak links can be identified and strengthened. In the initial stages of programme delivery, a thorough process evaluation can identify any potential weaknesses (and strengths). Routine monitoring of delivery and outcomes can ensure that the programme is consistently delivered per the original model, and to a high standard. Typically, facilitators have monitored programme implementation via self-reported checklists tailored to a specific parenting programme and designed by the programme developers. This approach has limitations as developer involvement has the potential to introduce subjective bias into the instrument design, and tailoring to specific content makes it difficult to compare fidelity across different parenting programmes. We propose a move towards the application of objective measures that can be applied generically across a range of parenting programmes with high levels of reliability and validity.

The purpose of the current study is therefore to describe a simple tool that was developed to address this need, the Parent Programme Implementation Checklist (PPIC, Bywater, 2011),and to explore its initial ability to reliably measure fidelity, when applied to assess the Incredible Years (IY) pre-school BASIC parenting programme (Webster-Stratton, 2010). Additional data from other parent programmes will be utilised in further testing of the tool in due course.

Defining Implementation Fidelity

Implementation fidelity is becoming increasingly important with efforts to scale evidence-based programmes and deliver results within mainstream services. Unfortunately, evidence indicates that when interventions are replicated in real-world contexts the outcomes often do not match those achieved in research settings (Alexander, Robbins & Sexton, 2000; Hutchings et al., 2007; Sexton & Turner, 2010). Subsequently, implementation research is important and should continue throughout scale-up to maintain the intervention's integrity and effectiveness,

and to ensure that programmes are not delivered by insufficiently trained staff with inadequate resources (Mihalic, Fagan, Irwin, Ballard, & Elliott, 2004; Moore et al., 2015). 'Implementation fidelity' is considered the degree of fit between the original programme and its application in each service setting. Five primary elements are suggested to comprise fidelity (Adherence, Exposure, Quality of Programme Delivery, Participant Responsiveness and, Programme Differentiation) and are all considered critical to the success of any parent programme (Mihalic et al., 2004), such as the IY programme as outlined below.

Adherence describes whether or not the programme's content and procedures were delivered as designed i.e. all core components delivered to appropriate population. In terms of operationalising this during IY programme delivery facilitators are encouraged to promote relationship-enhancing and discipline, or limit-setting strategies, emphasise the need for parents to learn the principles of the programme such as sensitivity or reinforcement and, enhance their knowledge of child development in terms of what their child's capabilities are at each developmental stage. This learning should be supported through the use of videotaped vignettes to prompt discussion and problem-solving amongst the group, role-play to practice and rehearse new skills, and homework to consolidate learning between sessions. Typically processes associated with programme adherence during delivery are monitored through the use of facilitator completed checklists however, the principles of adherence should be embedded in practice from the start by ensuring that delivery staff are properly in trained in the programme, and have access to appropriate ongoing support and supervision to minimise 'drift'.

Exposure describes whether or not the treatment 'dose' matches the original programme i.e. number and length of sessions. In terms of monitoring this during routine delivery of IY facilitators may complete checklists to record how much content of each session was delivered, as well as recording parent attendance each week to monitor programme 'dosage' i.e. the number of sessions parents attend. Such information is particularly useful when trying to establish the relationship between programme delivery and family outcomes.

Quality of programme delivery refers to whether the manner of delivery, the skill of facilitators in using the materials, techniques or methods is consistent with what is expected and prescribed by the programme. During IY programme delivery this element of implementation fidelity is operationalised through the use of a collaborative and reciprocal relationship between group leaders and parents, with the emphasis being that both parties have expertise. As a result, facilitators should use their skills to encourage parents to solicit their ideas and participate in personal goal setting. Moreover, facilitators should adapt the intervention to meet their parents own individual needs by spending more time on programme

content that parents in their group may need more support on. Self-report checklists completed by facilitators routinely monitor this aspect of implementation fidelity.

Participant responsiveness describes the extent to which the participant is involved in the activities and content of the programme i.e. contributes to group discussions. This often focuses on the degree to which parents feel empowered to find their own solutions, feel encouraged to help each other and build support networks. During IY, and other, programme delivery this aspect of implementation fidelity is often monitored via weekly and end of programme parent-reported evaluation forms.

Finally, *programme differentiation* identifies the unique or critical components of a programme that reliably differentiates it from others, or the comparison intervention. This typically refers to whether or not the core (or essential) programme sessions are being delivered as specified in the programme manual; these processes or content are commonly monitored through the use of weekly facilitator-completed checklists.

Carroll et al., (2007) suggests that these five individual elements of fidelity act as potential moderators of the relationship between interventions and their intended outcomes. Subsequently, the degree in which these elements are met during delivery affects how well the programme succeeds in achieving its goals of promoting change.

Implementation Fidelity within Programme Delivery

The National Institute for Clinical Excellence (NICE) guidelines (2017) for treating children and adolescents at-risk of, or diagnosed with, oppositional defiant disorder or conduct disorder recommends the use of psychosocial interventions such as group based parenting programmes as an alternative treatment to pharmacology. The recommendations stipulate that group-based interventions that are manualised and that involve parents should utilise behavioural or cognitive-behavioural approaches and subsequently draw on social learning theory principles (Bandura, 1977) in programme content and delivery i.e. modelling, rehearsal and parent feedback, to improve parenting skills. Programmes are suggested to be at their most efficacious if delivered to groups of 10 to 12 parents once a week for 90 to 120 minutes over the course of 10 to 16 sessions. Exemplars of psychosocial programmes, such as IY (Webster-Stratton, 2010) and Triple P (Sanders, Markie-Dadds, Tully & Bor, 2000), involve an interactive and collaborative learning format in which programme facilitators discuss and model key behavioural principles and parenting skills (e.g. play, praise, rewards, and discipline) to parents and caregivers, who then practise these skills in and outside of group sessions. Key components of the most effective programmes include: learning how and when to use positive parenting

skills; observation; modelling; behaviour rehearsal (e.g. role-play); discussion; homework assignments; using peer support, reframing unhelpful cognitive perceptions about their child or child-management; and, tackling barriers to attendance (Gardner, 2012; Hutchings, Gardner, & Lane, 2004). These features notably define the *Adherence*, *Quality* and *Participant Responsiveness* elements of implementation fidelity and are the most commonly assessed aspects of programme delivery as facilitators can self-monitor their own ability in attaining these goals to be effective (Hutchings et al., 2004).

Evidence indicates that greater fidelity to the model is linked to improved outcomes for participants, whilst results are weaker where implementation fidelity is poor (e.g. Blakely et al., 1987; Botvin, Baker, Filazzola & Botvin, 1990; Durlak & DuPre, 2008; Eames et al., 2009; 2010; Lee et al., 2008; Kam, Greenberg & Walls, 2003; Pentz et al., 1990; Rohrbach, Graham & Hansen, 1993). A handful of studies have also tested whether a causal relationship exists between programme fidelity and outcomes for children and families, however, this research is largely correlational and contradictory. For example, several studies report positive and significant relationships between fidelity and outcomes (e.g. Eames et al., 2009; 2010; Forgatch, Patterson & Gewitz, 2013), whilst others report mixed or no significant findings (Breitenstein et al., 2010; Hogue et al., 2008; Malti, Ribeaud & Eisner, 2011). Whilst there is variation in how fidelity has been defined, operationalised and measured across different studies, it is reasonable to assume that the outcomes drawn from any evidence-based parenting programme are dependent on facilitator skills and expertise. For example, even though a programme is manualised, it is a facilitator skill to be able to relate the content and attend to the needs of each specific group of parents within their local context, by drawing upon their skills and knowledge as a practitioner. As a result facilitator behaviour should be the focus of routine monitoring over the course of programme delivery to ensure that parents are provided with high quality supervision with the best chances of instigating behaviour change.

Current Methods for Assessing Fidelity

Many evidence-based parenting programmes have infrastructure to support the monitoring and promotion of fidelity, some more extensive than others. In addition, during initial evaluation there are a number of methodological practices that researchers can engage in to ensure that studies reliably test interventions as they would be delivered given optimal conditions in routine practice. Garbacz et al. (2014) reviewed the use of strategies to promote fidelity as reported in 65 research trials of evidence-based parent training programmes designed to reduce child and adolescent behavioural difficulties. Using the Intervention Fidelity Assessment Checklist (IFAC), a tool developed to aid consistency in the assessment of fidelity

promoting and monitoring strategies in evaluation studies of behaviour change interventions (Bellg et al., 2004), the authors demonstrated that 75% of the 65 included studies described the use of fidelity strategies as part of methodological practice (treatment design [programme differentiation], training providers [quality of programme delivery], delivery of treatment [adherence and exposure], receipt of treatment [participant responsiveness], and enactment of treatment skills [quality of programme delivery]), with only five (8%) reporting high adherence (>80%) to fidelity strategies across all five categories. These five studies included two reporting the IY BASIC programme (Fossum, Morch, Handegard, Drugli & Larsson, 2009; Reid, Webster-Stratton & Hammond, 2007), one reporting the Triple P programme (Morawska & Sanders, 2009), one study (Kazdin, Siegal & Bass, 1992) reported on Problem-Solving Skills Training and Parent Management Training (PSST and PMT) and one on Behavioural Parent Training (BPT for ADHD; Thompson et al., 2009). The findings from this review suggest that it is not always clear whether programme content is fully adhered to, even within research studies, and as with any self-report measure, subjective bias from the facilitators can influence the outcome (Green, Goldman & Salovey, 1993). In addition, there is often variable quality across different programmes in their monitoring and supporting of fidelity as part of routine practice. For example, some programmes insist that accreditation and ongoing supervision are essential to ensure effective programme delivery whilst others require initial training only.

In response to the limitations of self-report and problems with integrating routine fidelity monitoring into programme delivery for both practice and research, a handful of independent observational tools of programme fidelity have been developed. Such tools utilise a range of scoring methods i.e. rating scales, checklists or frequency counts of specified facilitators behaviours, and are typically developed for use with specific programmes, i.e. the Leader Observation Tool (LOT: Eames et al., 2009; 2010) for the IY parenting programmes, or the CAS-CBT (Bjaastad et al., 2015) for the Curious Cat programme. Whilst these tools evidence reliability and validity, they can be complex and require observers to undergo high intensity training in order to be fully competent with complex coding systems. They also only apply to the particular programme under observation, which can be problematic when service providers begin to embed a range of different programmes and are limited in time and money to independently assess delivery across a suite of interventions.

To circumnavigate these issues the Parent Programme Implementation Checklist (PPIC, Bywater, 2011), has been developed as a generic checklist to capture 'global' implementation of the core components of group-based parenting programmes. The main objective of the PPIC is to provide a simple method for conducting random checks on programme fidelity to prevent

programme 'drift'. The PPIC focuses specifically on the Adherence, Quality and Participant Responsiveness components of fidelity as these can be easily observed, and are less likely to be affected by subjective bias. The tool negates the need for the user to have detailed knowledge of facilitator process skills or in-depth programme content by providing a simple checklist for assessing quality of delivery by either a member of service delivery staff, or a researcher..

Aim

The purpose of the current study is to describe the development of the observational PPIC, and explore its initial psychometric properties and potential as a generic tool of assessing parenting programme delivery/implementation fidelity. In particular, we explore whether the items in PPIC are coherent and measure the same construct/s (internal consistency) and whether raters can consistently reach agreement, over time (intra-rater reliability) and with different observers (inter-rater reliability). Assessment of the achievable levels of inter- and intra-rater reliability are a pre-requisite for all observational tools during initial development and whilst being used out in the field. Moreover, these statistics are possibly the most important when considering the use of the PPIC as a routine tool to monitor implementation fidelity within routine practice. This is because inter- and intra-rater reliability provides an indication of how much consensus is achievable between different coders (for instance it is important that coders are using the tool in the same way so that a service can have confidence in the scores across their coders), or for one coder over a period of time following training in the use of the tool.

Method

Measure Development

The Parent Programme Implementation Checklist (PPIC; Bywater, 2011)

The PPIC is an 18-item tool that was originally developed to assess aspects of programme implementation fidelity by independent observation, for the purposes of providing an objective assessment of treatment integrity in three pragmatic randomised trials (see Bywater et al., 2009; Hutchings et al., 2007; Little et al., 2012; Morpeth et al., 2016). Initial work to develop the items focused specifically on reviewing the self-completion checklists of two evidence-based and widely delivered group-based parenting programmes i.e. IY preschool and school-age BASIC for parents of children aged 3 to 12 (Webster-Stratton, 2010; http://www.incredibleyears.com/resources/tm/) and Level 4 Triple P for parents of children from birth to 12 years (Sanders et al., 2000; http://www.triplep.net/glo-en/getting-started-

with-triple-p/implementing-triple-p/implementation-support/). The purpose of this activity was to ensure that the PPIC included similar items of fidelity that were routinely measured as part of programme delivery whilst establishing where additional items were needed to ensure that the five components of fidelity were addressed (see Table 1). During this initial scoping exercise several key elements of the parenting programmes were identified as not being recorded or monitored as part of regular programme checklists, for example, modeling of key parenting behaviours and role play. Consequently, common elements from effective programmes were mapped against one of the five fidelity components (Adherence, Exposure, Quality of Programme Delivery, Participant Responsiveness, and Programme Differentiation), and then quantified and operationalised along a 5-point Likert scale (ranging from 1 'not at all' to 5 'excellent') by defining distinct behaviours associated with each item at each level.

(Table 1 here)

Initial development feedback

In 2013 the developer sought qualitative feedback from 4 trained users of the PPIC and other experts in the field. The purpose of this exercise was to; 1) establish acceptability and user-friendliness of the tool, and, 2) identify any items that required further clarification or revision. Feedback and subsequent revision of the PPIC focused on two specific areas:

1) Clarifying and operationalising the definitions of individual items

2) Reconstructing the sub-scales to increase their construct and face validity

In terms of clarifying and operationalising the definitions of individual items the following revisions were made; A) one item relating to the facilitators use of questions were separated into two distinct items (open-ended questions and problem-solving) to highlight their individual value (now items 5 and 6). B) Definitions listed in the training manual for items relating specifically to questions, homework review and role-play were given more detail. Finally, C) additional description was added to several items within the tool itself to ensure that the PPIC captured the ability of the facilitators to respond to the parent's needs (items 2, 15 and, 16).

With regards to the construct and face validity of the subscales, feedback from trained users led to the re-classification of Exposure and Programme Differentiation as components of

Adherence (now items 16 and 17). Thus the components of fidelity assessed by the PPIC were reduced from five to three (*Adherence*, *Quality* and *Participant Responsiveness*), in addition to an overall *Total Score* (Table 2). Space for information on both exposure and dose has been maintained at the top of the coding sheet to provide information about the context of the session i.e. number of parents attending the session, and total length of the session.

Maximum scores for the three components of fidelity are as follows; range 15 for Participant Responsiveness, 35 for Quality and, 40 for Adherence. The maximum attainable Total Score is 90. Currently there are no cut-offs for this measure; that is, we cannot say if a score above or below a certain level yields good versus poor outcomes (this will be addressed in future PPIC work). However, the higher the total fidelity score the more effective each session/programme may be in achieving positive outcomes in families (Eames et al., 2009; 2010). In addition, there is no current agreement about what constitutes a good or acceptable level of fidelity; theoretically this may be programme specific and there is considerable variability in the published empirical literature ranging from 60-90% (e.g. Botvin, 2004; Mihalic et al., 2004). The newly revised PPIC can now be used to code programme sessions either in-vivo (i.e. live by regular service staff or research team), or using videotaped recordings of individual sessions which is less obtrusive. This paper reports on videotaped observations only. Irrespective of the specific mode of observation chosen (i.e. live or video), in line with current observation recommendations, fidelity assessments should be consistently applied to prevent any confounds in the data caused by switching between observation modes (Gridley, Bywater, & Hutchings, in press).

(Table 2 here)

Validation Sample

Twenty-five, 2-hour video-recordings collected (with parent consent) from 14 independent groups delivering the 12-session IY BASIC parent programme in 13 localities across England and Wales in 2004 to 2009-10 as part of two large-scale randomised controlled trials (1. Hutchings et al., 2007; Bywater, Hutchings, Daley, Eames, Tudor-Edwards, & Whitaker, 2009, and, 2. Little et al., 2012; Morpeth et al., 2016), provided data for the study. The 25 videos were taken during either session 2 ($n = 1$), 4 ($n = 4$), 5 ($n = 8$) or 8 ($n = 12$). The 25 videos were independently reviewed and coded by two primary coders (20 and 5 videos respectively) who had received the PPIC training and were knowledgeable, but not trained/accredited,) in the IY BASIC parent programme. A secondary coder rated 16 of the videos for inter-rater reliability checks and 6 of the 16 videos (37.5%) were subject to code re-code (intra-rater) checks. The

final sample of data presented in this paper relates only to the 16 videos taken from 16 individual sessions (session 4 n = 2; session 5 n = 7; session 8 n = 7) from 10 independent groups conducted in 9 localities as these were coded by both the same primary and secondary coders.

Validation Procedure

The training model comprises a half-day to one-day group training session led by a trained user (the lead author), with frequent refreshers to maintain reliability of coding within organisations. Training is supplemented by a detailed manual which outlines each item to be coded, its definition, and examples. As part of the training coders are encouraged to ask questions about applying the tool prior to viewing and coding 'training' videos of other group sessions from the same programme. Following viewing of each video clip, the trainer checks coding reliability, and resolves discrepancies through group discussion. All coders reached a pre-specified level of 70% inter-rater reliability with the lead author prior to coding independently. The four coders were all educated to Master's degree level. Over the course of six months (2013-2014) the primary and secondary coders independently rated each of the videos in a quiet room using a stopwatch to time the sessions. Data from each of the coders was then entered into an SPSS database for analysis purposes.

Analysis Plan

To assess the internal reliability of the 18 different items of the PPIC, in addition to the three subscales (Adherence, Quality and Participant Responsiveness) and the overall Total Score a series of Spearman Rank correlations for categorical data were used. This type of analysis is important during the initial stages of tool development in order to test the assumption that individual items are measuring the same construct/s, and therefore that the outcomes are meaningful. To assess how closely related the set of items that sit under each of the three subscales and overall total score categories were a series of Cronbach Alphas were calculated (internal consistency).This level of analysis is important at all stages of tool development and later use as an assessment tool to ensure that the individual items that comprise a sub-scale are indeed measuring the same concept and therefore provide meaningful data to assess implementation fidelity.

Assessment of inter-rater reliability was conducted by applying three different types of reliability analysis to the 16 videos coded by the primary and secondary coders; 1) percentage agreements, 2) Intra Class Correlations (ICC's) using a two-way mixed model with absolute agreement, and 3) a weighted Kappa for categorical data. Using each method of analysis inter-

rater reliability was assessed at the item by item level and for each subscale and Total Score.

Intra-rater reliability was conducted on 6 videos coded by the primary coder. Two types of analysis were applied at both the item by item level and the subscale level; percentage agreements and ICC's using a two-way mixed model with consistency.

For interpretation purposes all reliability statistics scores ranged between 0 and 1 with larger scores indicating greater agreement between coders.

Results
Internal reliability and consistency of the PPIC

Table 3 presents a series of correlations conducted to assess the internal reliability between each of the 18-items of the PPIC (Table 3) and their respective subscales. With the exception of three items (3, 14 and 16) all other remaining items demonstrated at least one significant correlation (at the $p < .05$ level) with one other PPIC item. Correlations for these significant items ranged from $r = .500$ to $.900$ indicating moderate to strong consistency between items. These findings suggest that the 18-items of the PPIC are at some level inter-related and measuring similar constructs.

(Table 3 here)

The internal consistency of the PPIC subscales (Adherence, Quality and Participant Responsiveness), as well as the overall Total Score were assessed using Cronbach alphas[2]. Analysis indicated that the Adherence subscale, which consists of eight items, demonstrated questionable levels of internal consistency $\alpha = .661$. The Quality subscale, which consists of seven items, demonstrated acceptable levels of internal consistency $\alpha = .780$, whilst Participant Responsiveness, which consists of only three items, demonstrated low internal consistency, $\alpha = .440$. The overall Total Score value for the PPIC provided a good level of internal consistency $\alpha = .818$. These findings suggest that the items that make up the four sub-scales of the PPIC are measuring the same construct, thereby providing some evidence for the composition of these scales.

Achievable Levels of Agreement between Different Coders
Percentage agreements

---

[2] For interpretation of alphas: $\alpha \geq .90$ Excellent, $\alpha \geq .80$ Good, $\alpha \geq .70$ Acceptable, $\alpha \geq .60$ Questionable, $\alpha \geq .50$ Poor, $\alpha < .50$ Unacceptable

Percentage agreements, a quick and easy method to determine coder agreement, between the primary and secondary coder indicated achievement to the minimal acceptable level of >70% for inter-rater reliability (Aspland & Gardner, 2001). Results indicated that the mean agreement achieved between coders across all 18 items was 70.62% (*SD* = 9.51). Agreement ranged between 54% and 88% suggesting that reasonable levels of agreement, as calculated using percentage agreements, could be achieved between two coders who received half-a-day training in using the PPIC and who were not necessarily experienced in observational methods. This suggests PPIC could be used easily by a variety of individuals or organisations.

Intra-Class Correlations (ICC's)

Table 4 presents the findings from a series of two-way mixed model ICC's[3] with absolute agreement, a more rigorous method of testing agreement amongst coders. At the item level ICC's ranged between -.025 and .864 indicating no or some agreement between coders with a large correlation. Only six of the 18 items (see Table 4) indicated statistically significant agreement (*p* < .05). These items indicated moderate levels of agreement between coders. Items 3, 4, 6, 7, 8, 10, 13, 14, 15, 16 and 17 demonstrated little or no agreement between primary and secondary coders and were therefore not statistically significant.

At the subscale and Total Score level ICC's ranged between .404 and .730 suggesting agreement between coders. Agreement between coders reached statistical significance (*p* < .05) across three of the four categories with a medium to large correlation. These results suggest reasonable levels of inter-rater reliability are achievable at the subscale level when calculated using ICC's but not at the item by item level.

(Table 4 here)

Weighted Kappa

Table 4 also presents the findings from the weighted Kappa[4] analysis, used because the Likert response scale of the PPIC is categorical in nature, and because Kappa is the most robust method for assessing agreement amongst coders. Overall the results replicated those found using the ICC's with one exception; an additional significant agreement was found between coders on item nine for off task behaviour. Kappa coefficients ranged from -.013 to .764 with the seven statistically significant items indicating agreement between the two coders with a poor to substantial effect. Items 3, 4, 6, 7, 8, 10, 13, 14, 15, 16 and 17 demonstrated little or no agreement between primary and secondary coders and were not statistically significant.

---

[3] For interpretation of ICC results: ≥.75 Excellent, ≥.60 Good, ≥ .40 Fair, < .40 Poor
[4] For interpretation of weighted Kappa values: ≥.81 Very good, ≥ .61 Good, ≥ .41 Moderate, ≥ .21 Fair, < .20 Poor

(Table 5 here)

Achievable Levels of Agreement for One Coder at Two-Time points

Percentage agreement

Assessment of the 6 videos subject to code re-code analysis by the same coder indicated reasonable achievable agreement across all 18 items. Percentage agreements for intra-rater reliability was slightly higher than that achieved for inter-rater analysis with a mean of 72.71% (*SD* = 6.65), agreement ranging between 60% and 79%. These findings suggest that reasonable levels of agreement could be achieved by the same coder when using the PPIC to code the same videos at two different time points.

ICC's

A series of two-way mixed model ICC's with consistency were conducted to assess intra-rater reliability using six videos that had been re-coded by the same primary coder. Table 5 presents the results. ICC's ranged from -.143 to .935 suggesting varying levels of achievable intra-rater agreement at the item by item level. Only four of the 18 items indicated statistically significant agreement, all with large correlations (see Table 5). Items 2, 3, 4, 5, 6, 7, 8, 9, 10, 13, 14, 15, 16 and 17 demonstrated little or no agreement and were therefore not statistically significant.

At the subscale level ICC's ranged between .176 and .939. Only the Quality subscale did not reach statistical significance. The remaining three subscales demonstrated intra-rater agreement with medium to large correlation. The findings suggest that intra-rater reliability using the PPIC at the item level is poor, however high levels can be achieved at the subscale and Total Score level. This is important as the sub-scales and total score values are most likely to be used as part of practice to monitor progress. Moreover, whilst many coders may be trained to use the PPIC it is important that each organisation has one lead coder i.e. the most experienced and reliable, who can conduct the majority of observations and ensure that other coders maintain their reliability levels over a period of time. Intra-rater reliability is therefore an important assessment to establish how stable a coder's score is over time.

Discussion

The current paper describes the initial development and assessment of the internal reliability and consistency, and achievable levels of inter- and intra-rater reliability of a generic

fidelity tool to assess implementation fidelity for group-based parenting programmes. The PPIC was developed to measure the principles of Adherence, Quality of delivery, and Participant Responsiveness for group-based parent programmes, and although the current paper addresses the usability of this tool with only one programme, IY BASIC, it is reasonable to expect that the fidelity items could be applied to other group-based parenting programmes too.

The results indicate adequate levels of internal reliability and consistency for the 18-individual PPIC items, three associated subscales and the overall Total Score. Achievable levels of inter and intra-rater reliability between coders were lower than expected at the item by item level, but met the recognised standards of reliability at the subscale and Total Score level. These findings indicate some promise of the PPIC to be used as an assessment tool of implementation fidelity for parenting programmes. However, caution is warranted if applying these results in routine practice to monitor implementation fidelity. Further work is required to refine the tool to ensure that it meets statistical standards for reliability and validity across a variety of different programmes. Furthermore, additional development is required to ensure that the tool is user friendly for a range of personnel who may not be familiar with observational methods i.e. non-research staff.

Levels of agreement between coders using the PPIC were poorest on items in the following conditions; 1) where there is a degree of ambiguity or subjective interpretation in the definitions that are to be quantified and applied (i.e. models problem solving questions, models acknowledgements, uses praise, off-task, encourages participation), 2) which may require a degree of observer subjectivity, or that may be lost due to poor film quality (i.e. use handouts smoothly, homework explained, video clips used, sum up important points from session), and 3) items which require explicit knowledge of the programme under study (i.e. key concepts covered, non-session content excluded). Previous research has suggested that in order for fidelity tools to be successful a comprehensive coding manual should be developed to support its implementation (Forgatch, Patterson, & DeGarmo, 2006). It is suggested that this manual should include information defining each program component, outline the procedures for scoring observations, and specify the rating scheme to be used (Forgatch, et al., 2006). Moreover, to maximize observer objectivity and reduce subjectivity, each point on a given item's scale should be anchored to quantify specific behaviours or practices (Mowbray, Holter, Teague & Bybee, 2003). Whilst the coders were provided with a half-day training it is possible that some of the inconsistency in their overall agreement at the item by item level may have resulted from their lack of experience in using observational methods. For example, some of the items of the PPIC require attention from coders across the whole video (models praise) and are

therefore based on number of occurrences or frequency counts whilst other items relate specifically to discrete behaviours that may only appear once during the course of the session i.e. explain homework. A coder with less experience of using observational methods may not easily grasp the difference between these types of items and subsequently further work is required to enhance the usability of the PPIC manual so that it can be accessed and understood relatively easily by professionals with limited expertise in observational methods. As a result, the current findings are important if the PPIC is to be integrated into routine practice as coders indicated that greater knowledge of the tool (i.e. training and coding maintenance) and programme content may be needed to sustain high levels of consistency and reliability. Consequently, further refinement of the tool via quantification and operationalisation of definitions laid out within the coding manual is needed if it is to function as an assessment tool for practitioners, group leaders, and possibly service managers, as part of routine practice, as well as other researchers.

Strengths

The main strength of this study is that there is a real need in research and in practice to develop tools that allow for the objective assessment of implementation and fidelity of parenting programmes without increasing costs or the time needed to train personnel in becoming reliable in applying the measure, or, indeed the actual parent programme being observed. As a result, the PPIC can be regarded as a much needed tool in parent programme research and programme fidelity assessment. Current tools can be time consuming and costly in terms of training and applying the tool, e.g. the LOT (Eames et al., 2009) was designed to observe a specific parent programme, and may require extensive knowledge of the programme content. This is the first study to assess whether the PPIC can reliably measure fidelity and the current findings suggest cautious optimism, particularly at the subscale and total score level. Whilst further refinement of the individual items is needed, in addition to further validation of the tool when used to assess other group-based parent programmes, or when used by non-research staff, the PPIC does show some promise of being able to reliably assess the fidelity of group-based programmes. This quick and easy to use measure does not eliminate the need for facilitators to access supervision during delivery, nor does it alter the need for working towards programme delivery accreditation (which may entail programme trainers or developers giving in-depth feedback following observations), where relevant. The strength lies in the fact that the PPIC can possibly be used across programmes, and can be used to identify great delivery, but

also identify when facilitators could benefit from additional training or more supervision to ensure positive outcomes for families.

Limitations

There are several limitations of the current study. Firstly, despite the intention to address the shortfalls of previous implementation self-complete checklists, or observational measures, by providing evidence for a tool that can be used across a range of different programmes, the current study explored implementation fidelity for only one parent programme i.e. the dataset used to assess the reliability of the PPIC relates to only one group-based parent programme (IY BASIC). There were two reasons for this; 1) from the original batch of videos collected during the two RCT's only four were taken during Triple P programme delivery. The other three Triple P videos were subject to technical issues at the programme delivery stage and were excluded from any assessment. 2) In order to ensure that the data that we had would be suitable to conduct inter-rater analysis for the purposes of this paper we restricted the remaining dataset ($n = 25$) to only those videos that had been coded by the same pair of coders ($n = 16$). As a result, the findings supporting the reliability of the PPIC are limited to only the IY BASIC intervention, and to a small set of videos that were deemed clean and audible (discussed in more detail below). Consequently, the findings cannot yet be generalised to other group sessions and therefore further study of the PPIC's ability to effectively measure implementation fidelity of other group-based programmes, and within usual service delivery, is required as its current ability to be used as a generic tool is aspirational.

A second limitation of the current study is that the data used to assess the reliability of the PPIC is drawn from a programme delivered within research settings as part of previous pragmatic community based RCT evaluations. Programme delivery in the context of research evaluation is known to be more adherent, and previous validation studies using other fidelity tools have often utilised larger samples of videotapes, reviewing at least two videos from each group delivery to capture the variation in delivery that would prompt different fidelity assessments. Due to technical issues with a minority of the videotapes (cameras were not turned on, or switched off half-way through the session) the number of videos eligible for use was reduced and the current data reflects only those that were deemed clean and audible. In routine practice the use of videotape technology is often a problem, and there is a question as to whether the current set of videotapes is likely to reflect the real world context of programme implementation. In order to mitigate such issues arising in real world delivery where programmes are videoed as usual i.e. for the benefits of supervision, future revisions of the tool

and its associated manual will need to include clear guidance on how facilitators should set up and position the camera during delivery in order to ensure that clean and audible video recordings are possible. As a general guide it is suggested that a minimum of two sessions captured at random points across programme delivery is recommended to enable fidelity scores to be calculated (Barber et al., 2006). We suggest, in our training manual, that the first and last sessions of programme delivery should be avoided from fidelity coding due to the (respectively) introductory and celebratory nature of these sessions but that the two chosen sessions per group should be filmed approximately ¼ and ¾ way through the programme to enable a better perspective of how programmes are being delivered over time, and because fidelity of programme delivery is not necessarily static.Moreover, in terms of applying the PPIC we recommend a half day to one day training, with frequent refreshers to maintain reliability of coding within organisations. The benefits of this manualised step by step approach to recording and coding session delivery includes; 1) aiding the process of built in supervision and accreditation for programme facilitators by ensuring that suitable videos are available, and, 2) ensuring that cameras are positioned in such a way that PPIC coders are able to rate all items which will in turn enhance the ability to continually monitor the maintenance of delivery at a fairly low cost.

The final limitation is that whilst the original PPIC set out to capture all five aspects of programme fidelity, based on feedback from fidelity experts the final revised tool only relates to three (Adherence, Quality of delivery and Participant Responsiveness). The previously included items relating to the components of Exposure (appropriate number of sessions) and Programme Differentiation (unique features of programmes) were identified as being best aligned with Adherence. Whilst outstanding items were suggested to be best captured in greater detail through other means i.e. weekly facilitator completed checklists and attendance logs, and not through randomly selected observations. Whilst this is a shortfall of the tool, it is acknowledged that these refinements have strengthened the tools reliability and face validity for these three components.

Future Directions and Conclusions

Work is underway to explore the feasibility of using the PPIC with other group-based parenting programmes. This work will explore achievable levels of reliability when using a larger sample of videos derived from several different programmes and their individual sessions. We intend to explore the content validity and structural validity of the PPIC, and liaise with programme developers and experts in fidelity research to revise the PPIC manual, training,

and coding sheets. Once the PPIC tool has been fully validated we intend to conduct a study to explore the reliability of using the PPIC live during session delivery versus using the PPIC from pre-recorded videotapes as we have done so with other observational tools (Gridley, Bywater & Hutchings, in press). It is hoped that findings from such a study would be useful and informative in instances where video-technology may not be available in practice yet assessment of programme fidelity is still required.

Results suggest that the half-day training yields reasonable levels of inter-rater reliability, to individuals not trained in the parent programmes they are observing. However, the tool is not yet systems tested and further work with the PPIC when used by non-research staff in practice-based settings, who may have little knowledge of observational methods, is required before we can be confident that the training programme and tool could be used by a variety of individuals across organisations.

In conclusion, developing tools that serve to measure implementation fidelity of parent programmes within real world settings is an important and challenging area of work, namely to justify the expense of delivery in the face of increasing cuts to services and to ensure that families receiving these services stand a good chance of benefiting from its content. To our knowledge a similar tool for generic assessment of quality across multiple parent programmes has not been successfully developed. The PPIC is in its initial stages of development and the current study suggests that it has the potential to make a real world contribution to an area where routine monitoring is important to ensure that quality standards of programme delivery are upheld and maintained. However, the findings should be taken with caution as the current study is exploratory in nature and has a number of limitations which may have contributed to the findings. More importantly, the PPIC is an observational tool which in real world settings would most typically be applied during live delivery. For validation purposes, video is clearly the most optimal choice because we need different coders to view it. As a result, we acknowledge that there is a difference here between how the tool is used as part of its initial development and the future use of the tool as it is being implemented in services following validation. Our initial intentions of the PPIC was to address all five components of fidelity, yet two components were subsumed within the 'adherence' component, and other parts of exposure and differentiation were captured in the session's information at the top of the form (i.e. length of programme, whether the correct session content was delivered in the appropriate week, and number of parents attending the session to establish if the group is a 'viable' group for learning and discussion purposes).

References

Alexander, J. F., Robbins, M. S., & Sexton, T. L. (2000). Family-based interventions with older, at-risk youth: From promise to proof to practice. *Journal of Primary Prevention*, *21*(2), 185-205. doi: 10.1023/A:1007031219209

Aspland, H., & Gardner, F. (2003). Observational measures of parent-child interaction: An introductory review. *Child & Adolescent Mental Health*, *8*(3), 136-143.

Axford, N., Bywater, T., Blower, S., Berry, V., Baker, V., & Morpeth, L. (2017). *Critical factors in the successful implementation of evidence-based parenting programmes: fidelity, adaptation and promoting quality.* In What works in child protection: An evidence-based approach to assessment and intervention in care proceedings.

Bandura, A. (1977) Social Learning Theory. General Learning Corporation, NY.

Barber, J. P., Gallop, R., Crits-Christoph, P., Frank, A., Thase, M. E., Weiss, R. D., & Beth Connolly Gibbons, M. (2006). The role of therapist adherence, therapist competence, and alliance in predicting outcome of individual drug counselling: Results from the National Institute Drug Abuse Collaborative Cocaine Treatment Study. *Psychotherapy Research*, *16*(02), 229-240. doi: 10.1080/10503300500288951

Barlow, J., Smailagic, N., Ferriter, M., Bennett, C., & Jones, H. (2010). Group-based parent-training programmes for improving emotional and behavioural adjustment in children from birth to three years old. *Cochrane Database of Systematic Reviews*, *3*. doi: 10.1002/14651858.CD003680.pub2.

Barlow, J., Smailagic, N., Huband, N., Roloff, V., & Bennett, C. (2012). Group-based parenting programmes for improving parental psychosocial health. *Cochrane Systematic Reviews (6).* doi: 10.1002/14651858.CD.002020.pub.3

Barlow, J., Smailagic, N., Huband, N., Roloff, V., & Bennett, C. (2014). Group-based parent training programmes for improving parental psychosocial health. *The Cochrane Library*. doi: 10.1002/14651858.CD002020.pub4.

Bellg, A. J., Borrelli, B., Resnick, B., Hecht, J., Minicucci, D. S., Ory, M., Ogedegbe, G., Orwig, D., Ernst, D., & Czajkowski, S. (2004). Enhancing treatment fidelity in health behaviour change studies: best practices and recommendations from the NIH Behaviour Change Consortium. *Health Psychology, 23,* 443-451. doi: 10.1037/0278-6133.23.5.443

Bjaastad, J. F., Haugland, B. S. M., Fjermestadm K. W., Torsheim, T., Havik, O. E., Heiervang, E. R. & Ost, L. G. (2015). Competence and adherence scale for cognitive behavioural therapy (CAS-CBT) for anxiety disorders in youth: psychometric properties. *Psychological Assessment, online first.* doi: 10.1037/pas0000230

Blakely, C. H., Mayer, J. P., Gottschalk, R. G., Schmitt, N., Davidson, W. S., Roitman, D. B., & Emshoff, J. G. (1987). The fidelity-adaptation debate: implications for the implementation of public sector social programs. *American Journal of Community Psychology, 15,* 53-268. doi: 10.1007/BF00922697

Botvin, G. J., Baker, E., Filazzola, A. D., & Botvin, E. M. (1990). A cognitive behavioural approach to substance abuse prevention: one year follow up. *Addictive Behaviours, 15,* 47-63. doi: 10.1016/0306-4603(90)90006-J

Botvin, G. J. (2004). Advancing prevention science and practice: Challenges, critical issues, and future directions. *Prevention Science*, *5*(1), 69-72.

Breitenstein, S. M., Gross, D., Garvey, C., Hill., C., Fogg, L., & Resnick, B. (2010). Implementation fidelity in community-based interventions. *Research in Nurse Health, 33,* 164-173. doi: 10.1002/nur.20373

Bywater, T. Hutchings, J., Daley, D., Eames, C., Tudor-Edwards, R. & Whitaker, C. (2009). A pragmatic randomised control trial of a parenting intervention in sure start services for children at risk of developing conduct disorder; Long term follow-up. *British Journal of Psychiatry, 195, 318-324.*

Bywater, T. (2011). *The Parent Programme Implementation Checklist Training Manual.* University of York, Unpublished manuscript

Bywater, T. (2012). Developing Rigorous Programme Evaluation. Chapter 3 in Barbara Kelly & Daniel F. Perkins (Eds) *The Cambridge Handbook of Implementation Science for Education: Psychology in Education – How to Promote Evidence Based Programmes and Practices*. Cambridge University Press.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, *2*(1), 40. doi**:** 10.1186/1748-5908-2-40

Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: a review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology, 41,* 327-350. doi: 10.1007/s10464-008-9165-0

Eames, C. et al. (2009) Treatment fidelity as a predictor of behaviour change in parents attending group-based parent-training. *Child: Care, Health & Development 35,* 603–612. doi: 10.1111/j.1365-2214.2009.00975.x

Eames, C., Daley, D., Hutchings, J., Whitaker, C. J., Bywater, T., Jones, K., & Hughes, J. C. (2010). The impact of group leaders' behaviour on parents' acquisition of key parenting skills during parent training. *Behaviour Research & Therapy*, *48*(12), 1221-1226. doi: 10.1016/j.brat.2010.07.011

Forgatch, M. S., Patterson, G. R., & Degarmo, D. S. (2006). Evaluating fidelity: predictive validity for a measure of competent adherence to the Oregon model of parent management training. *Behaviour Therapy, 36,* 3-13. doi: 10.1016/S0005-7894(05)80049-8

Forgatch, M. S., Patterson, G. R., & Gewitz, A. H. (2013). Looking forward: the promise of widespread implementation of parent training programs. *Perspectives in Psychological Science, 8,* 682-694. doi: 10.1177/1745691613503478

Fossum, S., Morch, W. T., Handegård, B. H., Drugli, M. B., & Larsson, B. O. (2009). Parent training for young Norwegian children with ODD and CD problems: Predictors and mediators of treatment outcome. *Scandinavian Journal of Psychology*, *50*(2), 173-181. doi: 10.1111/j.1467-9450.2008.00700.x

Furlong, M. et al. (2012) Behavioural/cognitive-behavioural group based parenting interventions for children age 3–12 with early onset conduct problems. *Cochrane Systematic Reviews, 15.* doi: 10.1002/14651858.CD008225

Garbacz, L. L., Brown, D. M., Spee, G. A., Polo, A. J., & Budd, K. S. (2014). Establishing treatment fidelity in evidence-based parent training programes for externalising disorders in children. *Clinical Child Family Psychology Review, 17,* 230-247. doi: 10.1007/s10567-014-0166-2

Gardner, F. (2012). Parenting interventions: Effectiveness across contexts and cultures. 15th Annual Conference of Helping Families Change Conference HFCC), Glasgow. Retrieved from http://www.pfsc.uq.edu.au/about/hfcc/pdf/Day%202%20-%20Thursday%20-%209th%20of%20February/Plenary%20Sessions/Parenting%20Interventions%20Effectiveness%20A cross%20Contexts%20and%20Cultures%20-%20Frances%20Gardner.pdf

Green, D.P., Goldman, S.L., & Salovey, P. (1993). Measurement error masks bipolarity in affect ratings. *Journal of Personality & Social Psychology, 64,* 1029–1041. doi:10.1037/0022-3514.64.6.1029

Gridley, N., Bywater, T., & Hutchings, J. (in press). Comparing live versus video observation: using the dyadic parent interaction coding scheme (DPICS) as an example.

Hogue, A., Henderson, C. E., Dauber, S., Barajas, P. C., Fried, A., & Liddle, H. A. (2008). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behaviour problems. *Journal of Consulting & Clinical Psychology, 76,* 544-555. doi: 10.1037/0022-006X.76.4.544

Hutchings, J., Bywater, T., Daley, D., Gardner, F., Whitaker, C., Jones, K., Eames, C., & Edwards, R. T. (2007). Parenting intervention in Sure Start services for children at risk of developing conduct disorder: pragmatic randomised controlled trial. *British Medical Journal*, *334*(7595), 678. doi: 10.1136/bmj.39126.620799.55

Hutchings, J., Gardner, F., & Lane, E. (2004). Making evidence-based interventions work. *Support from the start: working with young children and their families to reduce the risks of crime and anti-social behaviour* (in Norwich: Department for Education and Skills (Research report 524).

Kam, C., Greenberg, M. T., & Walls, C. T. (2003). Examining the role of implementation quality in school-based prevention using the PATHS curriculum. *Prevention Science, 4,* 55-63. doi: 10.1023/A:1021786811186

Kazdin, A. E., Siegel, T. C., & Bass, D. (1992). Cognitive problem solving skills training and parent management-training in the treatment of antisocial-behaviour in children. *Journal of Consulting & Clinical Psychology, 60,* 733–747. doi:10.1037//0022- 006x.60.5.733.

Lee, C. Y., August, G. J., Realmuto, G. M., Horowitz, J. L., Bloomquist, M. L., & Klimes-Dougan, B. (2008). Fidelity at a distance: Assessing implementation fidelity of the Early Risers Prevention Program in a going-to scale intervention trial. *Prevention Science. 9*, 215–229. doi: 10.1007/s11121-008-0097-6

Little, M., Berry, V., Morpeth, L., Blower, S., Axford, N., Taylor, R., Bywater, T., Lehtonen, M., & Tobin, K. (2012). The impact of three evidence-based programmes delivered in public systems in Birmingham, UK. *International Journal of Conflict & Violence*, *6*(2), 260-272.

Malti, T., Ribeaud, D., & Eisner, M. P. (2011). The effectiveness of two universal preventive interventions in reducing children's externalizing behaviour: a cluster randomized controlled trial. *Journal of Clinical Child & Adolescent Psychology*, *40*(5), 677-692. doi: 10.1080/15374416.2011.597084

Mihalic, S., Fagan, A., Irwin, K., Ballard, D., & Elliott, D. (2004). *Blueprints for Violence Prevention.* Boulder, CO: University of Colorado, Center for the Study and Prevention of Violence.

Moore, G., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wight, D., Baird, J. (2008). *Process evaluation of complex interventions: Medical Research Council guidance*. MRC Population Health Science Research Network, London, 2014. Retrieved from http://decipher.uk.net/wp-content/uploads/2014/11/MRC-PHSRN-Process-evaluation-guidance.pdf

Moore, G. F., Audrey, S., Barker, M., Bond, L., Bonell, C., Hardeman, W., Moore, L., O'Cathain, A., Tinati, T., Wright, D., & Baird, J. (2015). Process evaluation of complex interventions: Medical Research Council guidance. *British Medical Journal*, *350*, h1258. doi: 10.1136/bmj.h1258

Morawska, A., & Sanders, M. (2009). An evaluation of a behavioural parenting intervention for parents of gifted children. *Behaviour Research & Therapy*, *47*(6), 463-470. doi: 10.1016/j.brat.2009.02.008

Morpeth, L., Blower, S., Tobin, K., Taylor, R. S., Bywater, R. J, Edwards, R. T., Axford, N., Lehtonen, M., Jone, C., Berry, V. (2016). The effectiveness of the Incredible Years pre-school parenting programme in the UK: a randomised controlled trial. *Child Care in Practice, online first edition.*

Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, *24*(3), 315-340.

NICE (2017). *Antisocial behaviour and conduct disorders in children and young people: recognition and management.* Clinical Guideline 158. Retrieved from https://www.nice.org.uk/guidance/cg158/resources/antisocial-behaviour-and-conduct-disorders-in-children-and-young-people-recognition-and-management-pdf-35109638019781

Pentz, M. A., Trebow, E. A., Hansen, W. B., MacKinnon, D. P., Dwyer, J. H., Johnson, C. A., Flay, B. R., Daneils, S., & Cormack, C. (1990). Effects of program implementation on adolescent drug use behaviour; the Midwestern prevention project (MPP). *Evaluation Review, 14*, 264-289. doi: 10.1177/0193841X9001400303

Reid, M. J., Webster-Stratton, C., & Hammond, M. (2007). Enhancing a classroom social competence and problem-solving curriculum by offering parent training to families of moderate-to high-risk elementary school children. *Journal of Clinical Child & Adolescent Psychology*, *36*(4), 605-620. doi: 10.1080/15374410701662741

Rohrbach, L. A., Graham, J. W., & Hansen, W. B. (1993). Diffusion of a school-based substance abuse prevention program: predictors of program implementation. *Preventive Medicine, 22,* 237-260. doi: 10.1006/pmed.1993.1020

Sanders, M., Markie-Dadds, C., Tully, L. A., & Bor, W. (2000). The Triple P Positive Parenting Programme: a comparison of enhanced, standard and self-directed behavioural family intervention for parents of children with early onset conduct problems. *Journal of Consulting & Clinical Psychology, 68*, 624-640. doi: 10.1037/0022-006X.68.4.624.

Sexton, T., & Turner, C. W. (2010). The effectiveness of functional family therapy for youth with behavioural problems in a community practice setting. *Journal of Family Psychology*, *24*(3), 339. doi: 10.1037/a0019406

Thompson, M. J. J., Laver-Bradbury, C., Ayres, M., Le Poidevin, E., Mead, S., Dodds, C., et al. (2009). A small-scale randomized controlled trial of the revised new forest parenting programme for pre-schoolers with attention deficit hyperactivity disorder. *European Child & Adolescent Psychiatry, 18,* 605–616. doi:10. 1007/s00787-009-0020-0.

Webster-Stratton, C. (2010). *The Incredible Years Parenting Programme User's Manual.* Unpublished manuscript.

Table 1.

Individual components of fidelity, how they are defined, operationalized and measured during routine programme delivery, in addition to how outstanding items were addressed in development of the PPIC

Component of Fidelity	Definition	Operationalised within Parenting Programme Delivery	Routine Methods for Monitoring	Outstanding Items Addressed by PPIC
Adherence	Assessing whether the programme is being delivered as it was designed, with all the core components, to the appropriate population, with staff trained to the appropriate standard, with the right protocols, techniques and materials and in the prescribed locations or contexts.	• Relationship-enhancing and discipline or limit-setting strategies. • Emphasis on parents learning 'principles', such as the need for sensitivity and reinforcement. • Child development knowledge and awareness of children's capabilities. • Videotaped vignettes prompt discussion and problem-solving. • Role-play rehearsal of new skills. • Homework with practice assignments.	• Facilitator completed checklists • Ensuring all delivery staff have had appropriate training and experience with access to support/supervision to minimise drift • Ensuring 'standardised' complete sets of prepared programme materials, are available to all groups	• Appropriate videotapes for the session are used • Role-play is included to reinforce learning • Agenda is explained • Homework is reviewed from previous week • Homework for following week is explained • Weekly session content is covered • Non session-specific content is excluded • Summing up important points

Table 1.

Individual components of fidelity, how they are defined, operationalized and measured during routine programme delivery, in addition to how outstanding items were addressed in development of the PPIC

Component of Fidelity	Definition	Operationalised within Parenting Programme Delivery	Routine Methods for Monitoring	Outstanding Items Addressed by PPIC
		• Parents are encouraged to keep records of their practice at home, and to set their own weekly goals • Parents receive weekly feedback from group facilitators.		

Table 1.

Individual components of fidelity, how they are defined, operationalized and measured during routine programme delivery, in addition to how outstanding items were addressed in development of the PPIC

Component of Fidelity	Definition	Operationalised within Parenting Programme Delivery	Routine Methods for Monitoring	Outstanding Items Addressed by PPIC
Exposure	Whether the treatment 'dose' (e.g. the number of parenting sessions in a course, and their frequency and length) matches the original programme.		• Weekly facilitator completed checklists • 'Dosage' also refers to number of sessions attended per parent related to outcome - this is assessed by attendance records and outcome measures.	• Length of session is appropriate

Table 1.

Individual components of fidelity, how they are defined, operationalized and measured during routine programme delivery, in addition to how outstanding items were addressed in development of the PPIC

Component of Fidelity	Definition	Operationalised within Parenting Programme Delivery	Routine Methods for Monitoring	Outstanding Items Addressed by PPIC
Quality of Programme Delivery	The manner of delivery, the skill of facilitators in using the techniques, or methods, their enthusiasm, preparedness and attitude.	• A collaborative, reciprocal relationship, which assumes that the facilitators and the parents both have expertise. Facilitators solicit parents' ideas and parents participate in goal-setting and are encouraged to adapt the intervention to meet their own individual needs.	• Weekly facilitator completed checklists • Peer & self-facilitator completed checklists, e.g. in weeks 4 & 8 for IY	• Inclusion of all parents • Model open-ended and problem-solving questions • Model acknowledgment and praise • Prevent side-tracking by parents • Prepared materials for ease of delivery

Table 1.

Individual components of fidelity, how they are defined, operationalized and measured during routine programme delivery, in addition to how outstanding items were addressed in development of the PPIC

Component of Fidelity	Definition	Operationalised within Parenting Programme Delivery	Routine Methods for Monitoring	Outstanding Items Addressed by PPIC
Participant Responsiveness	The extent to which the participant is involved in the activities and content of the programme.	• Parents are empowered to find their own solutions. • Parents are encouraged to help each other, reducing isolation and building support networks, by, for example, making calls to one another during the course • Group facilitators phone parents during the course, and contact parents who miss any sessions.	• Parent completed weekly evaluation forms e.g. IY • Parent completed end of programme evaluation forms e.g. IY and TP	• Parents participate in role-play • Each parent contributes to discussion elements • Each parent completes homework
Programme Differentiation	Identifies the unique features of different components of programmes that are reliably differentiated from one another.	• Course content delivered within predefined sessions	• Weekly facilitator completed checklists	• Correct session is delivered in the time slot, i.e. are sessions delivered in correct order?

Table 2.
Item by item descriptions of the PPIC and their associated subscales.

Component of fidelity	Description & guidance	Item
Adherence		
Does the facilitator present and explain the 'agenda'?	Agenda to be presented verbally, written on flipchart or projected	1
Does the facilitator review/discuss homework from previous week?	Facilitator comments and offers feedback to parents on completed homework to ensure understanding and gauge progress	2
Do facilitators encourage 'role-play' congruent with the session's key concepts (or as a solution to a homework problem from the previous week)?	The role-play should be congruent with the session's key concepts. Role-play is defined as 'practising verbal or nonverbal behaviour'. Facilitators should encourage parents to try different techniques, strategies, words to see how it feels/works	12
Are video clips congruent with the session's key concepts and used appropriately?	Video clips should relate to the key concepts for the session	13
Does the facilitator sum up important points relating to key concepts from the session?	Facilitators should reiterate the important main points from the session to encourage learning, ideally during and at session end	14
Is the homework for the following week explained?	Facilitator should give parents clear guidance and instructions for next week's 'homework', which may include specific practice	15
Are weekly session key concepts covered?	See summary of key concepts: IY wk 4 relates to praise, IY wk 8 relates to effective limit setting, TP wk 2 relates to promoting child development including praise and modelling, TP wk 4 relates to planning ahead with rules and consequences	16

Component of fidelity	Description & guidance	Item
Does the facilitator only include content and key concepts from this session (last week's content may be reviewed as appropriate)?	This differs from going off task as it includes incorporating content from other sessions (or even other programmes). Reviewing the previous session does not count as non-session content as this is my be a requirement of the programmes	17
Quality of programme delivery		
Does the facilitator use programme materials/handouts smoothly?	Facilitator runs session with all handouts available in the correct order. Shows preparedness. Smooth handing of materials with no time delay reduces possibility of distraction and time delays	3
Does the facilitator encourage all parents to participate?	Facilitator tries to include each participant during the session in some way, e.g. asks questions to individuals, encourages role-play	4
Does the facilitator use or model 'open-ended' questions?	An open-ended question is a question that cannot be answered by 'yes' or 'no' it encourages a more detailed response from the parents, for example, 'What are your thoughts on the child's behaviour in that video-clip?'. Problem solving questions are also open-ended questions and encourage critical thinking and can include the problem definition, solution or consequence and can be used to identify own or others' feelings. Examples include, 'what would you do if….?' What do you think will happen if..?', 'How do you think it made him feel when he was praised?'	5
Does the facilitator use or model 'problem-solving' questions?	Problem solving questions are also open-ended questions and encourage critical thinking and can include the problem definition, solution or consequence and can	6

Component of fidelity	Description & guidance	Item
	be used to identify own or others' feelings. Examples include, 'what would you do if....?' What do you think will happen if..?', 'How do you think it made him feel when he was praised?'	
Does the facilitator model 'acknowledgment'?	Facilitator acknowledges parent comments/responses by; responding yes/nodding, an acknowledgment is a brief verbal response to the verbalisation or behaviour that contains no manifest content other than a simple yes or no response to a question, or that communicates a recognition of something the parent has said or done, with no descriptive content, e.g. Uh uh, Sure, OK, etc.	7
Does the facilitator model praise?	Labelled/unlabelled praise – well done/well done for completing your homework, Labelled praise is any specific verbalisation that expresses a favourable judgment upon an activity or product	8
Does the facilitator prevent side-tracking or 'off-task' behaviour?	Does not allow an individual, or group discussion, to go off-task for longer than 5 minutes at a time. Facilitator is able to pull back group on-task within this timeframe. Watch is needed note time at first sign of off-task behaviour and when back on task. For example, a parent discussing child management issues during a holiday, and the conversation then turns to discussion around holidays in general. Video clips can be a source of side-tracking - do the group focus on the concept that the video is highlighting rather than focussing on unimportant issues such as home environment etc?	9
Participant responsiveness		

Component of fidelity	Description & guidance	Item
Does each parent contribute freely to discussion elements?	Discussion is important and can empower parents as it offers a chance to share successes as well as an opportunity to problem solve together. Are all parents willing to join in discussions?	10
Do parents participate in role-play (verbal or nonverbal practice)?	Role-play is defined as 'practising verbal or nonverbal behaviour' and can be difficult for parents to be involved in initially. Once parents participate they gain new insight in learning different strategies, and how it feels to be a parent or child in each strategy. A skilled facilitator will be able to get parents involved in role-play. Rehearsal of techniques during the sessions will enable parents to use the techniques more easily at home and encourage behaviour change. There may not be enough time for all parents to participate, it is important that at least some do.	11
Do parents spontaneously ask questions?	Parents' involvement, interest, and confidence can manifest in spontaneous questions to the facilitator or each other.	18

Table 3.
Spearman's correlations as a measure of internal consistency item by item ($N = 16$)

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	Ad	Qual	Res	Total
1	1.000	0.395	-0.051	0.393	-0.039	-0.048	0.630**	0.158	0.500*	0.550*	0.708**	0.681**	0.231	0.001	0.344	0.022	0.609*	0.018	0.824**	0.421	0.654**	0.769**
2		1.000	-0.234	0.495	0.411	0.376	0.578*	0.558*	0.314	0.109	0.211	0.322	0.053	-0.317	0.472	-0.167	0.219	-0.101	0.446	0.580*	0.205	0.522*
3			1.000	0.327	-0.108	-0.020	0.155	0.204	0.073	-0.192	0.009	0.034	0.223	0.055	-0.095	-0.095	-0.206	0.304	-0.132	0.323	0.125	0.183
4				1.000	0.409	0.266	0.756**	0.485	0.517*	0.143	0.103	0.248	0.314	-0.286	0.334	0.180	0.255	0.306	0.389	0.780**	0.279	0.602*
5					1.000	0.336	0.537*	0.509*	0.373	-0.368	-0.233	-0.002	0.313	-0.315	-0.072	0.401	-0.103	0.045	0.143	0.607*	-0.166	0.264
6						1.000	0.273	0.329	0.149	-0.026	-0.224	-0.182	0.356	-0.055	0.183	0.000	0.245	0.335	0.091	0.535*	0.052	0.238
7							1.000	0.539*	0.679**	0.061	0.399	0.425	0.550*	-0.054	0.361	0.248	0.418	0.348	0.718**	0.848**	0.514*	0.856**
8								1.000	0.312	-0.302	0.134	0.286	0.203	-0.387	0.096	0.096	-0.194	0.055	0.193	0.718**	0.171	0.436
9									1.000	0.169	0.357	0.415	0.427	-0.362	0.029	0.085	0.302	0.254	0.426	0.698**	0.390	0.563*
10										1.000	0.155	0.120	0.181	-0.204	-0.098	-0.308	0.659**	-0.040	0.254	-0.029	0.287	0.143

11	1.000	0.900**	0.206	0.217	0.438	0.265	0.265	0.276	0.802***	0.176	0.884***	0.731***
12		1.000	0.128	-0.031	0.315	0.270	0.111	0.245	0.732***	0.315	0.807***	0.737***
13			1.000	0.118	-0.069	0.309	0.469	0.586*	0.432	0.541*	0.495	0.533*
14				1.000	0.478	0.119	0.335	0.291	0.281	-0.267	0.230	0.137
15					1.000	0.107	0.417	0.076	0.604*	0.234	0.371	0.537*
16						1.000	-0.102	0.410	0.376	0.098	0.310	0.294
17							1.000	0.144	0.617*	0.273	0.421	0.495
18								1.000	0.301	0.301	0.613*	0.424

*** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$

Table 4.
Means, standard deviations, two way mixed model Intraclass Correlations (ICC) with absolute agreement and weighted Kappa coefficients to assess inter-rater reliability across each of the PPIC items (N = 16)

Item	*Primary Coder Mean (SD)*	*Secondary Coder Mean (SD)*	*ICC*	*p*	Kappa	*p*
1	3.19 (1.42)	2.94 (1.00)	0.455	0.035*	0.353	0.016*
2	3.75 (1.06)	3.63 (0.96)	0.628	0.004**	0.543	0.001***
3	3.19 (0.54)	3.81 (0.83)	0.156	0.209	0.059	0.289
4	4.00 (0.63)	3.69 (0.60)	0.322	0.084	0.133	0.212
5	4.13 (0.50)	4.25 (0.58)	0.571	0.008**	0.500	0.005**
6	3.44 (0.89)	3.25 (0.58)	0.151	0.286	0.242	0.068
7	4.25 (0.68)	3.87 (0.81)	0.167	0.246	0.111	0.252
8	3.19 (0.65)	3.13 (0.50)	0.331	0.106	0.211	0.121
9	3.88 (0.88)	3.75 (0.78)	0.540	0.014*	0.273	0.049*
10	4.19 (0.75)	4.06 (0.57)	-0.029	0.543	-0.013	0.531
11	3.13 (1.67)	3.06 (1.53)	0.864	0.000***	0.677	0.000***
12	3.06 (1.44)	3.19 (1.47)	0.821	0.000***	0.764	0.000***
13	3.69 (0.79)	4.19 (0.66)	0.205	0.173	0.150	0.153
14	2.31 (1.14)	2.69 (0.79)	-0.030	0.546	-0.075	0.674
15	4.13 (0.62)	4.06 (0.57)	-0.025	0.536	0.222	0.111

16	3.13 (0.62)	3.50 (0.82)	-0.119	0.686	-0.143	0.826
17	3.75 (0.86)	3.94 (0.68)	-0.145	0.703	-0.061	0.656
18	3.25 (1.12)	2.56 (0.89)	0.574	0.001**	0.329	0.005**
Adherence	27.00 (4.53)	28.13 (2.83)	0.404	0.053	-	-
Quality	26.06 (3.21)	25.75 (2.93)	0.466	0.034*	-	-
Responsiveness	10.56 (2.55)	9.69 (1.99)	0.730	0.000***	-	-
Overall Score	63.62 (8.57)	63.56 (5.56)	0.663	0.002**	-	-

*** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$

Table 5.

Means, standard deviations and two way mixed model Intraclass Correlations (ICC) with consistency to assess intra-rater reliability across each of the PPIC items ($N = 6$)

Item	*Primary Coder Time 1* *Mean (SD)*	*Primary Coder Time 2* *Mean (SD)*	*ICC*	*p*
1	2.67 (1.51)	3.00 (1.27)	0.828	0.011*
2	4.17 (0.98)	4.50 (0.84)	0.600	0.077
3	3.67 (0.82)	3.33 (0.52)	-0.143	0.620
4	4.00 (0.89)	3.83 (0.41)	0.414	0.178
5	4.33 (0.52)	4.50 (0.55)	0.000	0.500
6	3.00 (1.10)	2.83 (1.47)	0.238	0.304
7	4.50 (0.84)	4.33 (0.82)	0.585	0.084
8	3.67 (0.52)	3.83 (0.75)	0.320	0.242
9	4.17 (1.33)	4.00 (0.63)	0.369	0.208
10	3.83 (0.98)	4.33 (0.52)	0.108	0.409
11	3.33 (1.03)	3.50 (1.23)	0.935	0.001**
12	3.33 (1.03)	3.33 (1.03)	1.000	-
13	3.67 (0.82)	3.67 (0.82)	-0.800	0.985
14	2.67 (1.03)	2.33 (1.03)	0.125	0.395
15	4.33 (0.52)	3.83 (0.75)	0.640	0.061
16	3.33 (0.82)	3.33 (0.82)	0.571	0.090

17	3.67 (1.21)	3.83 (0.98)	0.110	0.408
18	3.33 (0.52)	3.17 (1.17)	0.918	0.002**
Adherence	27.83 (4.45)	27.83 (2.32)	0.698	0.041*
Quality	27.33 (4.68)	26.67 (3.39)	0.176	0.353
Responsiveness	10.50 (2.43)	11.00 (2.37)	0.939	0.001**
Overall Score	65.67 (10.67)	65.50 (5.99)	0.700	0.040*

*** $p \leq 0.001$, ** $p \leq 0.01$, * $p \leq 0.05$

Appendix A: PPIC Tool

Please email tracey.bywater@york.ac.uk to access the latest version of the PPIC and request permission to use.

PARENT PROGRAMME IMPLEMENTATION CHECKLIST (PPIC) VFeb2015: To assess the degree of adherence to the delivery model, quality of facilitator skill, and parent responsiveness when delivering group format parent groups.

Name of person completing this checklist as **primary coder/secondary coder** (circle as appropriate): …………………..…………………………………...

Date completed: …………………………………….. Name and type of Parent programme:………………………Session number, e.g. 2 (of 12):………………………

Centre/area:………………………………… Date session was delivered:……………………Number of parents attending this session:…………....Optimum number:……

Time session began and ended: Start time……………..Finish time…………Total timed length of session (minus break time)…………. Is this within 10% of expected time **Y/N** (circle)

ITEM	1 = not at all	2 = poor	3 = satisfactory	4 = good	5 = excellent	AD	QU	PR
1. Does the facilitator present and explain the agenda?	Not presented either visually nor verbally	Presented visually with no verbal explanation	Presented verbally with no visual aid to refer to throughout the session	Presented both visually with verbal description but facilitator does not check for parent understanding of content	Presented both visually and verbally with detail and facilitator checks for parent understanding of content, e.g. asks if any questions/input			
2. Does the facilitator review homework from previous week and give feedback?	No review or acknowledgement of homework, or effort, by parents	Reviewed homework with *some* parents but rarely gave feedback	Reviewed and gave feedback to *most* parents, e.g. by problem-solving parents' homework difficulties.	Reviewed with *most* parents, gave detailed responses including problem-solving, and used parent experiences to highlight key principles.	Reviewed all parents' homework in a sensitive way, asked for clarification where necessary –combination of 3 & 4 with *all* parents			

ITEM	1 = not at all	2 = poor	3 = satisfactory	4 = good	5 = excellent	AD	QU	PR
3. Does the facilitator use programme materials/handouts smoothly?	Failed to use any programme materials	Lack of preparedness, poorly organized e.g. missing materials, wrong paperwork/slides	Uses all programme materials but not fluidly, e.g. hesitantly, slowly, too rushed	Uses all programme materials in a proficient manner	High level of skill demonstrated when using materials and slides, uses materials in a timely, sleek fashion with confidence			
4. Does the facilitator encourage all parents to participate?	Makes no effort to build rapport or encourage participation	Does not notice or encourage the quieter or more nervous, less enthusiastic group members	Makes some attempt to encourage the *majority* of parents to participate	Makes some attempt to encourage *all* parents to participate	Constantly encourages all parents to participate by referring to each parent individually and noticing when a parent has not contributed and treats each parent as equally important and valued. Creates a feeling of safety and atmosphere of parent empowerment			
5. Does the facilitator model 'open-ended' questions?	Does not use open-ended questions	Uses open-ended questions unsuccessfully, i.e. does not give time for response	Rarely uses open-ended questions, but does give time to respond	Sometimes uses open-ended questions and gives time to respond	Frequent use of open-ended questions to facilitate discussion and gives opportunity to respond and also acknowledges parental responses			
6. Does the facilitator model 'problem-solving' questions?	Does not use problem-solving questions	Uses problem-solving questions unsuccessfully, i.e. does not give time for response	Rarely uses problem-solving questions, but does give time to respond	Sometimes uses problem-solving questions and gives time to respond	Frequent use of problem-solving questions to facilitate discussion and gives opportunity to respond and also acknowledges parental responses			

ITEM	1 = not at all	2 = poor	3 = satisfactory	4 = good	5 = excellent	AD	QU	PR
7. Does the facilitator model 'acknowledgment'	Does not use acknowledgement	Uses verbal acknowledgement inappropriately e.g. before parent has completed what they are saying, suggesting not being an effective listener	Rarely uses acknowledgement, either verbal or physical	Sometimes uses acknowledgement - verbal or physical	Frequent use of acknowledgement, both verbal and physical (e.g. nodding)			
8. Does the facilitator model 'praise'	Does not use praise	Uses only unlabeled praise, e.g. 'well done', 'great'	Uses unlabeled praise a lot more than labeled praise	Uses equal proportions of labeled and unlabeled praise	More frequent use of labeled praise, e.g. 'you have done a great job with your homework this week'			
9. Does the facilitator prevent side-tracking or 'off-task' behaviour?	Easily and frequently taken off-task for over 5mins, makes no attempt to get back on-task	Goes off-task easily and frequently, but makes unsuccessful attempts to get back on-task within 5 mins	When off-task facilitator is sometimes successful in getting group back on-task within 5 mins	Rarely goes off-task over 5 mins, can easily re-focus to on-task content	Excellent leader skills and checks individuals and group immediately when going off-task, maintains focus on session content.			
10. Does each parent contribute freely to discussion elements?	Lack of contribution from any parent	Only a *few* (minority) of parents contributed but were unenthusiastic, or had to be drawn in to a response. The majority made no response.	A *few* (minority) of parents contributed enthusiastically and spontaneously	The *majority* of parents contributed enthusiastically and spontaneously	*All* parents contributed enthusiastically and spontaneously, i.e. without having to be encouraged or prompted to participate			
11. Do parents participate in role-play? Role-play refers to either practicing what to say, or do in various contexts.	No-one participated/it was not offered	Only a *few* (minority) of parents contributed when invited, and they were unenthusiastic. The majority did not participate	A *few* (minority) of parents participated enthusiastically when invited to participate in role-play	The *majority* of parents that were invited contributed enthusiastically	*All* parents that were invited to participate contributed enthusiastically			

ITEM	1 = not at all	2 = poor	3 = satisfactory	4 = good	5 = excellent	AD	QU	PR
12. Do facilitators encourage role-play congruent with the session's key concepts (or as a solution to a homework problem from the previous week)? Role-play refers to either practicing what to say, or do in various contexts.	No – role-play not offered or encouraged	Facilitator is not confident in encouraging role-play/practice, and is unclear on how it relates to the key principles, fails to engage parents in any role play	Facilitator encourages a *few* (minority) of parents to participate in at least one role play/practice congruent with the session	Facilitator is successful in encouraging the *majority* of parents to participate in at least one role-play/practice at some point in the session	Facilitator skillfully encourages *all* parents to participate in several spontaneous role-plays/practices during the session and makes clear the relation between the role-play and the key principles, and asked how it felt afterwards.			
13. Are video clips congruent with the session's key concepts and used appropriately?	No – no clips used	Facilitator knowledge of clips is poor, e.g. shows clips that are incongruent to the session's key concepts, or appears unsure of how to use effectively in relation to topic	Facilitator shows congruent clips somewhat successfully but may use either too many or too few clips to enable meaningful discussion	Shows congruent clips and encourages discussion, but may not refer to parents' personal goals or learning principles	Facilitator skillfully uses congruent clips to spark discussion, and refers to parents' personal goals or learning principles relating to the clips, does not let the discussion of the clip to go on too long			
14. Does the facilitator sum up important points relating to key concepts from session?	No summing up at all	Attempts to (verbally or visually) sum up key concept points, but does not do so successfully, e.g. summarises only a minority of key points in an inconsistent manner	Briefly (either verbally or visually) sums up all key points made either during the session, or at the end, but not at both time points	Sums up key points, both during the session and at the end both verbally and visually	Sums up key points both verbally and visually, both during the session and at the end, and also checks for parental understanding			

ITEM	1 = not at all	2 = poor	3 = satisfactory	4 = good	5 = excellent	AD	QU	PR
15. Is the homework for the following week explained?	No - not at all	Yes, but very poorly, e.g. facilitator demonstrating lack of knowledge/clarity of what homework is about, does not check for parental understanding or fails to get everyone to understand the homework	Explained homework but room for improvement e.g. explained too briefly or in rushed manner at the end of the session, not checked parents' understanding of the homework, parents may ask for clarification	Homework clearly explained, but parental understanding not checked, parents may ask for clarification	Aims and objectives of homework explained clearly and concisely, as is the relationship of homework with the sessions concepts, parents' personal goals may be reiterated, parent understanding of homework is checked until facilitator is happy that everyone understands			
16. Are weekly session key concepts covered?	No – none covered	Not all covered and those that are not covered well at all, e.g. half the session spent on one concept with inability to direct the session appropriately	Not all are covered, but those that are covered well	Yes, all are covered but sometimes too much or too little time spent on particular concepts	Yes, all are covered skillfully with the facilitator tailoring the concepts to parents' needs and spending more time on those concepts most needed			
17. Does the facilitator only include content and key concepts from this session (last week's content may be reviewed as appropriate)?	No – content from another session or programme is heavily included	Facilitator uses some content from another session or programme and appears unsure of what content should be included in the session	Facilitator briefly uses content from another session, e.g. if failed to cover all concepts in last week's session they may be brought in here	Yes, facilitator only includes content from this session, but may not cover all in any depth	Yes, excellent adherence to session and programme content. No additional content included, keeps to timely delivery of session (no time to include other content)			
18. Do parents spontaneously ask questions?	No – not at all	Yes, but rarely and unenthusiastically	Yes, sometimes, but only a *minority* of parents ask questions	Yes, the *majority* of parents appear comfortable to ask questions spontaneously	Yes, *all* parents show an interest and enthusiasm for learning, from the facilitator and each other, and frequently ask questions			
DOMAIN TOTALS								

ITEM	1 = not at all	2 = poor	3 = satisfactory	4 = good	5 = excellent	AD	QU	PR
GRAND TOTAL (min 18, max 90)								
Calculate % score by dividing total score by 90 and multiplying by 100						%		