The Family Links Nurturing Programme

Findings and recommendations from a strategic review and development project

EXTENDED SUMMARY 3:

Development and testing of an Overarching Impact Measure

Deborah Ghate

The Colebrooke Centre for Evidence and Implementation

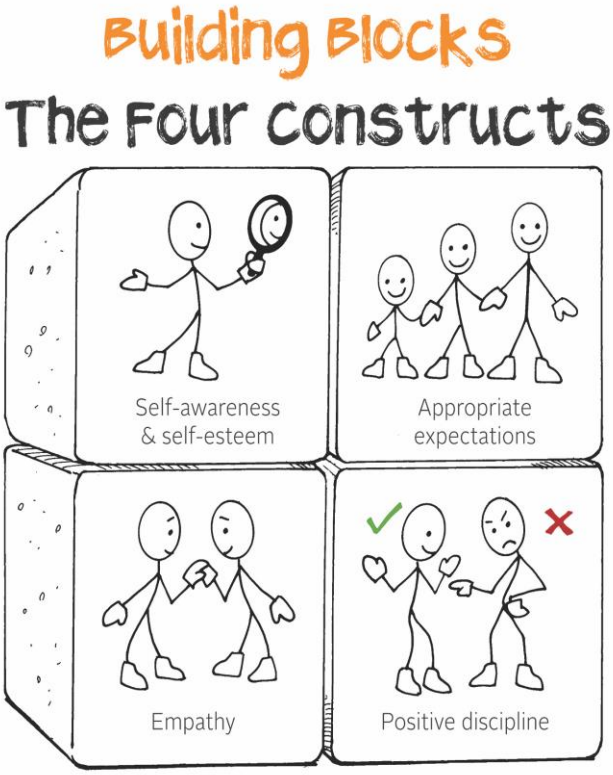
August 2015

This is a summary of key findings, conclusions and recommendations on work to develop a single overarching measure of impact for the Family Links Nurturing Programme (FLNP). The work formed part of an 18-month strategic review and development project carried out by the Colebrooke Centre for Evidence and Implementation for Family Links during the period Autumn 2013 to Spring 2015. A full report on the entire project was prepared for Family Links: for enquiries about that report, please contact: research@familylinks.org.uk

How to cite this summary:
Ghate D (2015) The Family Links Nurturing Programme - Findings and recommendations from a strategic review and development project Extended summary 3: Developing and Testing of an Overarching Impact Measure Oxford: Family Links

Background to the Family Links Nurturing Programme

Family Links was established as a registered charity in 1997. Operating out of two central offices (one in Oxford in the South East of England and one in Hull in the North East, which we refer to collectively as 'Head Office' in the report), the Family Links Nurturing Programme (FLNP) has now been delivered in the UK for over fifteen years and although the number of local authorities delivering the Programme at any one time varies, over 50 local authorities and voluntary organisations around the country in England, Wales, and Northern Ireland have the capacity to deliver it. A highly popular programme with commissioners and providers alike, it is well-loved by practitioners and parents and enjoys strong local commitment. FLNP is one of several different parenting programmes developed by Family Links and forms the 'core' or mainstream programme on which other variants for specific groups are based, including antenatal parents, parents in prison, Muslim parents, and parents of children with special needs. Mainstream FLNP is a 'broad spectrum' parenting support programme, not a child behaviour management or behaviour modification programme, and much of its content deals with general principles of healthy relationships in families and with aspects of parenting confidence and self-efficacy. As such, research (our own, and that of others) shows that parents find the content relevant across a wide range of child age ranges. This is considered a strength by its commissioners, who use it both as a universal intervention, and for selective prevention when it is offered to parents with particular needs for support. Local areas therefore refer parents with children of all ages, although with an emphasis on pre-school and primary school-aged children.



Based on '*four constructs*' that are the thematic building blocks around which content is delivered (self-awareness and self-esteem; appropriate expectations; empathy; positive discipline) the Programme locates its roots in a family of programmes known in the USA as the '*Nurturing Parenting Programs*'.

These were first developed in the 1970s and are now delivered in several countries, but predominantly in the USA (Bavolek, 2000). The Nurturing Parenting Programs are based on social learning principles, and most are described as child maltreatment prevention interventions. The Family Links Nurturing Programme in the UK is different, in that it explicitly does <u>not</u> describe itself as a child

maltreatment prevention programme, nor does it generally work with statutory referrals. It is offered in community-based settings open to the whole community, often in Sure Start Children's Centres, and although parenting support and other workers may (and substantially do) recommend the programme to parents they work with, in the majority of cases it is attended by parents on an entirely voluntary basis.

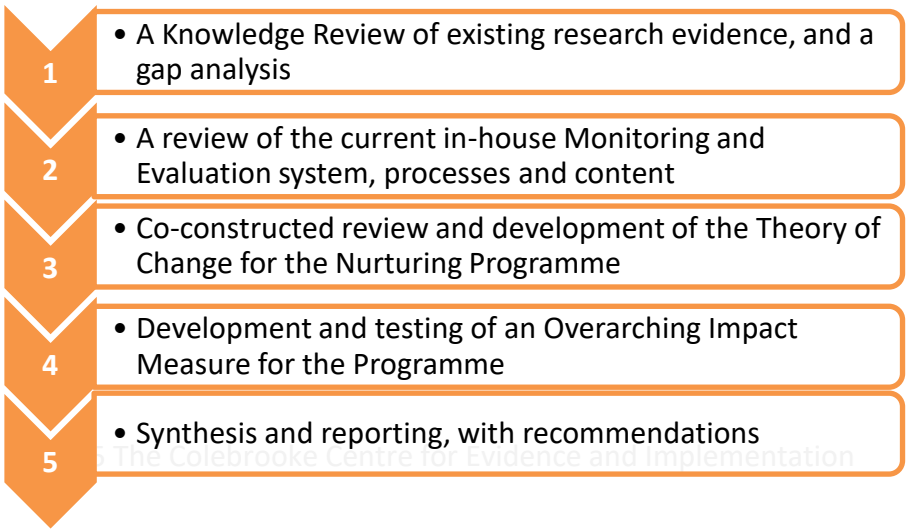
At local level, the Programme is delivered over a ten week period, sometimes with a prior week for introductions and preparation, in mixed groups usually intended to contain not more than ten parents, by trained facilitators, known by the Programme as Parent Group Leaders, PGLs). PGLS are not employed by Family Links, but the providing local authority or by other organisations to whom local authorities have contracted for the purpose of delivering the Programme. All PGLs have been trained in the Programme by Family Links, having undergone a four-day minimum training and approval process. There are two PGLs, per group. PGLs use a structured curriculum that is set out in the Parent Group Leaders' Handbook and in a book for parents, *The Parenting Puzzle: How to get the best out of family life* (Hunt, with Mountford, 2003). All parents receive their own copy of the book, and are encouraged to refer to it throughout the course.

The strategic review and development project

In 2013 Family Links commissioned the Colebrooke Centre to design a strategically-focused project to review and synthesise lessons from prior research on the Programme, and to carry out new work to review its foundations and identify ways to improve. As a specialist implementation analysis and improvement support centre, The Colebrooke Centre's approach uses an *implementation lens* (Fixsen et al., 2005) and draws on theory, frameworks and methods from intervention science and especially implementation and improvement science to inform and shape our work. Our approach, methods and many of our tools are innovative and unique in their application in children's services in the UK at present.

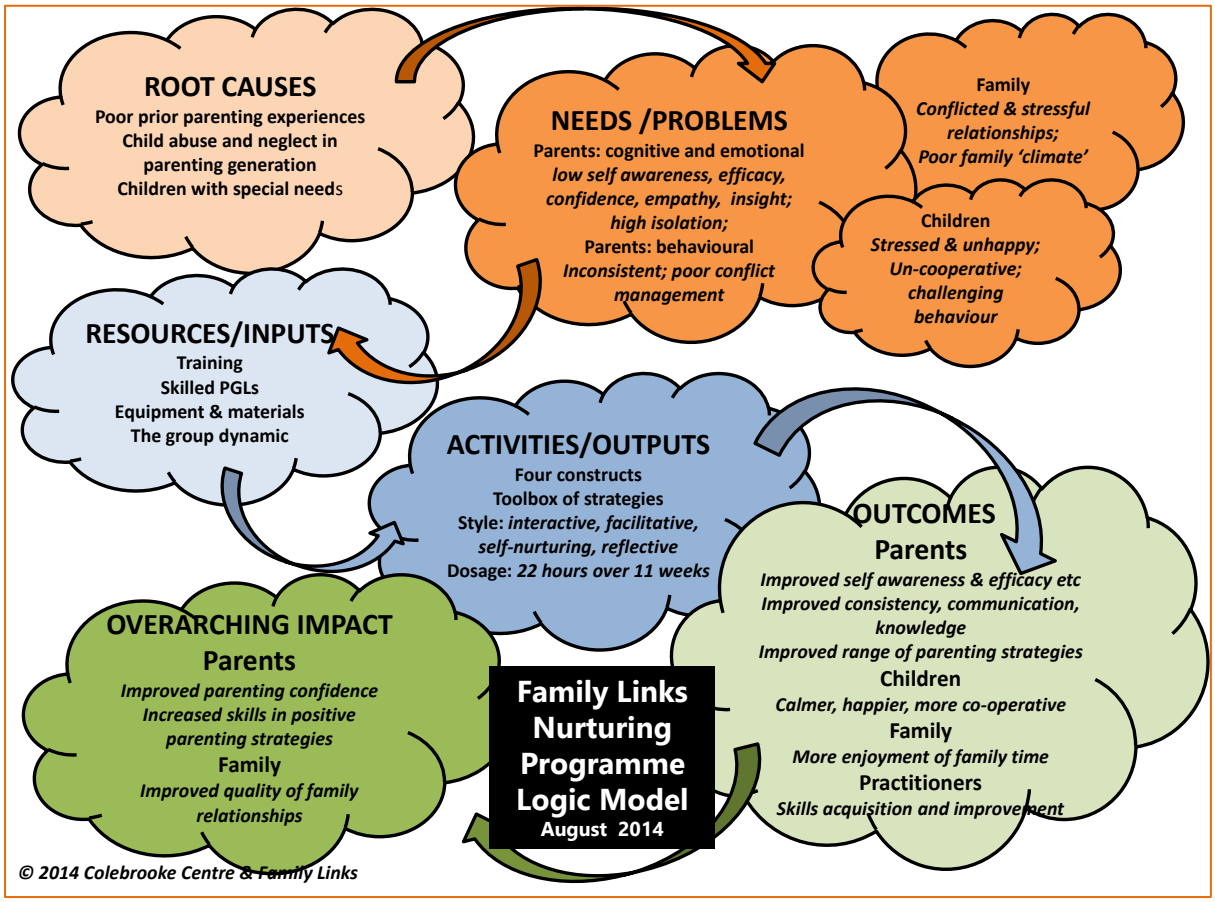
For this project, we designed a five stage approach as follows:

Figure 1 – the five-stage structure of the Strategic Review and Development Project



Below we summarise the key findings and implications from **Stage Four** of the work, a process to develop and test alternative measures suitable to be used to measure the overarching impact of the programme in future research and evaluation. This summary should be read in conjunction with Summaries (1) and (2) – the general summary of the whole project, and the summary of Stage Three, which reviewed the Programme's theory of change (Figure 2).

Figure 2 High level Logic Model (2) for the Family Links Nurturing Programme revised 2014



Development and Testing of an Overarching Impact Measure for the Family Links Nurturing Programme

Background and overview

The last empirical stage of our work for Family Links was to develop and field-test a small number of alternative measures potentially useful as an **overarching impact measure** for the FLNP. It is important to stress that the purpose of this stage was to develop and test some feasible measures, using appropriately robust and replicable methods, not to evaluate the effectiveness of FLNP. That stage, of using the measures 'at scale' to determine if FLNP 'works' is yet to come, and the results described below should be viewed in that light.

The concept of an overarching impact measure

The concept of an overarching impact measure (OIM) was first developed and described by the author and others in a collaborative project published in 2013 (see Moran and Ghate, 2013) and related publications[1]. In that work, we discussed the common problems faced in evaluations of social programmes in which complex designs and multiple, multi-layered outcome measurement schema not infrequently produce results that are difficult to interpret, whilst at the same time overlooking the higher-level purpose of the intervention. We believe that overarching impact measures may have important and useful value for social programmes, insofar as they attempt to cut through some of the complexity, and create a helicopter mechanism for 'seeing the wood for the trees'. A key feature of our approach is that we make a firm distinction between *outcomes* and *impact*, where impact describes a 'higher order' level of benefit based on cumulative change in multiple contributing outcomes.

In essence, an impact measure differs from an outcome measure in its degree of specificity. Outcome measures typically try to capture, in fine-grained ways, specific types of change, and so tend to be person-specific and domain-specific. An impact measure on the other hand is intended to provide a short, preferably single-scale measure of what we might call the 'essence' of a programme. It is intended to capture the 'main thing' (or things) that the programme designer hopes will change for all or most participants as a result of

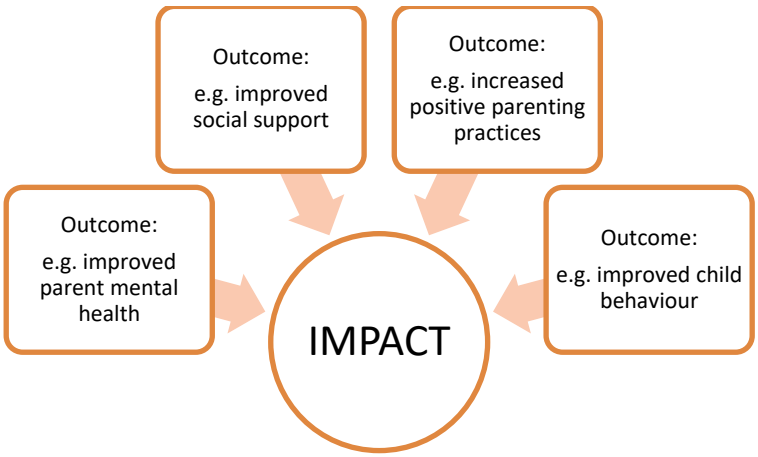
Distinction between impact and outcomes, from Moran and Ghate (2013)

Impact is the term used to describe the 'sum of the parts' of multiple outcomes: that is, the overarching, 'high-level' effect of a service that may be targeting a number of subsidiary outcomes. *Outcomes* relate to changes over time in specific 'domains' of functioning or life circumstances (e.g, parents' methods of discipline, children's behaviour, mothers' mental health, and so on).

[1] Available at http://www.cevi.org.uk/docs/Impact_Measure_Report.pdf

taking part. An impact measure is therefore more like an aggregate measure; see Figure 3.

Figure 3 A model for the relationship between outcomes and impact (from Moran and Ghate, 2013)



In order to be robust, an overarching impact measure should logically proceed from the specific outcomes the intervention intends to achieve (ie, it should not be chosen on a purely 'aspirational' basis). Our conception of such a measure is that it should be simple (though not simplistic) and brief. It should not be used to replace comprehensive and detailed evaluation of a programme using fine-grained exploration of outcomes (which aim to tell us 'what works for whom in what circumstances'), but can be used to give commissioners, providers and practitioners a trustworthy, low-cost, easy to administer, *broad-brush overview* of whether a programme is achieving positive change of the intended sort for the people who complete it. A good impact measure should therefore be appropriate to measure change for all or most participants in an intervention (i.e., not specialised to particular presenting needs or circumstances). It should also be acceptable to participants, and should be recognised by practitioners and other programme content experts as a reasonable test of their success. Above all, an OIM should closely reflect the agreed theory of change and be appropriate to the implementation model for the programme. Finally, it should be appropriate for programmes to use for self-evaluation or for low-cost independent evaluation, and so be easy to administer and analyse.

In Moran and Ghate (2013) we discuss in more detail the theory underpinning our concept of an OIM, the methods we have previously used for developing and testing these sorts of measures, and (an important '*a priori*' question) the broader methodological enquiries undertaken to confirm that simple, single-scale measures can indeed be robust, reliable and fit for the purposes intended. Readers who are interested in knowing more about the thinking that led to the original development work are invited to consult the report on that work (http://www.cevi.org.uk/docs/Impact_Measure_Report.pdf).

Methods for developing an overarching impact measure for FLNP

To develop an OIM for the FLNP, the prior work on the theory of change in Stage Three of the project (see www.FamilyLinks.org.uk) was taken on to a further stage. We extrapolated from the consensus on specific outcomes to discussion and agreement about the overarching (or underlying), fundamental purpose of the programme. We asked the content experts: What is the *raison d'être* of the FLNP? *If nothing else changed* for families who had been exposed to the FLNP, what one or two key differences would you nevertheless expect to have made?

Three key dimensions for overarching impact were eventually distilled, within which positive changes for participating families would signify the success of the programme:

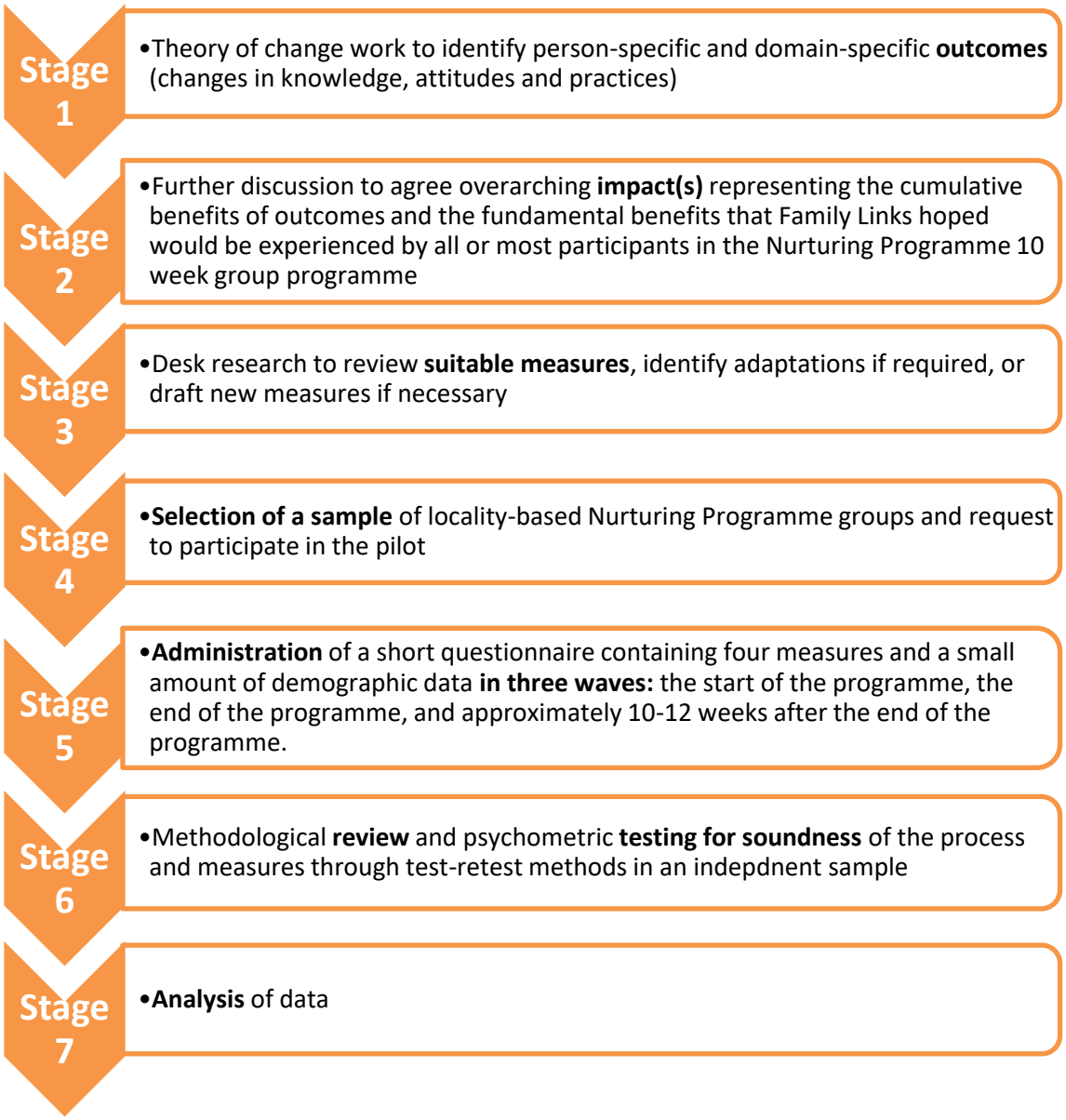
(a) the **quality of family relationships** (or alternatively **family climate** or home 'atmosphere' as a proxy);
(b) **parenting practices** as a proxy for parenting strategies (which are complex and multi-dimensional and therefore not usually measured);
(c) **parents' sense of coping and self-efficacy**.

After a process to identify and develop a selection of potential measures of these constructs, we then proceeded to field test a small number of these in a national sample of 30 Nurturing Programme groups in a 'pilot' exercise. With the assistance of FLNP Parent Group Leaders and their colleagues in the five UK regions in which the FLNP is delivered, we tested the selected measures, by inviting participating parents to self-complete their responses. The pilot test took place over a period of approximately 6 months, in three waves of 'repeated measures' data collection, starting in Autumn 2014. This process allowed us to check that the measures could be completed by parents in real Nurturing Programme settings, as well as whether the measures were sensitive to change over time.

In a parallel process in different groups, we also conducted a 'test-retest' exercise to collect data on the internal 'psychometric' properties of the measures (that is, how reliable and stable they were over time, and how valid). We also sought feedback by pro-forma and by more in-depth telephone interviews with PGLs to obtain their perspectives on the process and the qualities of the measures. Below we describe the methods used in this project, and the findings of the pilot testing phase.

The work to develop an overarching impact measure comprised of seven stages, as shown in the Figure 4 below:

Figure 4 Summary method for development of an Overarching Impact Measure for FLNP



Four scaled measures were selected /adapted for testing. Much more detail about the process, including detailed rationales for any adaptions made and samples of the scales can be found in the full report on the project (available at research@familylinks.org.uk).

Creating the shortlist of measures to be tested

Ideally of course, researchers prefer to use existing, pre-validated and standardised measures wherever possible. In fact there are in existence a huge number of potential measures on each of the three agreed impact areas that could in theory be considered. Many of them have been pre-tested to some degree. We did not conduct a systematic review of all potential measures for reasons of time and budget, and it is always possible that we missed some candidate measures. In addition, new measures become available all the time. For both of these reasons, and because the theory of change for the programme is likely to continue to evolve and be refined over time, it will be important to keep the measures under review.

This noted, the key drawback of most of the measures we reviewed was **substantive.** Measures are almost always developed for specific interventions, which generally rest on particular foundations with underlying principles, theories of change, content, target populations and target outcomes that are specific to that intervention or programme. This made them of questionable validity as measures of the FLNP.

In addition, because most measures are *outcome* rather than *impact* measures, their **design or structure** may be inappropriate for use as an OIM. Many are very long and complex or time-consuming to complete, and not couched in plain English suited to a UK population with sometimes limited literacy (as is typical of FLNP service users) [2].

Finally some have only been infrequently used in the field, and/or provide little information about their **reliability and validity**.

Therefore, we approached the task of developing an OIM shortlist as a involving a combination of review of existing measures, consideration of adaptations, development of new measures, and (critically) independent psychometric testing of a set of alternatives before fixing on a recommended measure.

Below we summarise the four measures that were selected /adapted for testing. Much more detail about the process, including detailed rationales for any adaptions made, can be found in the full report on the project (available at research@familylinks.org.uk).

[2] These caveats even apply, for example, to the AAPI-2 measure developed for evaluation of the Nurturing Programmes in the USA, (https://www.assessingparenting.com/assessment/aapi) . in which the FLNP has its roots. In the case of AAPI-2, although it does test constructs that are closely related to the FLNP, it is developed and has only been tested for use in the target populations for Nurturing Programmes in the USA (welfare populations – or as we would call them social care populations - that is, families on the edge of care or involved with statutory social work services for child safeguarding reasons, and clearly a higher risk group than the typical FLNP group). Also, at 40 items, it is too long to use as a brief impact measure.

Family relationships and family 'climate'

(1) The quality of family relationships

Improving the quality of family relationships was felt to be a primary overarching objective of the FLNP, flowing logically from successful achievement of the specific outcomes for parents and children detailed in the logic model. After a review of competing possibilities, we (rather heavily) adapted the widely used 12–item *McMaster Family Assessment Device* of global family functioning sub-scale ('*FAD-12'*; Epstein, Baldwin and Bishop, 1983) as the best available a measure of global family/household functioning for our purposes. Its advantages include that it has been widely used, especially in the USA, is well-validated, and many studies document its association with other measures of family and individual functioning.

The eventual measure, "*How Things Are in My Family"*, was an 11 item scale, and used a five point Likert-style response format. It is shown in the Appendix to the full report together with the original FAD-12 for comparison.

(2) Family climate

As an alternative and proxy construct, we agreed that a measure of 'family 'climate' or atmosphere in the home would also be an acceptable indicator of the quality of family life, reflecting the work done throughout the 10 week NP course to encourage parents to work towards equilibrium in what is referred to as 'the temperature' at home: hot (angry relationships, conflicts, high levels of emotional outbursts); cold (distant or un-empathic relationships, lack of mutual involvement and shared activities); or warm (the ideal; well-regulated and mutually supportive relationships).

We did not however find a satisfactory pre-existing measure of 'family climate' that mapped well to the FLNP, so we designed a bespoke scale for this purpose, drawing from items used in scales in regular use in large scale surveys of parenting in the UK and US, and adding new items that were assesses to particularly reflect FLNP's content. The eventual measure consisted of seven items, using a seven-point style Likert scale. The measure developed, "The Family Links *Family Climate Scale*" is shown in the Appendix to the full report.

(3) Parenting practices

The second impact area that we chose to focus on was movement towards more positive parenting strategies by parents completing the programme. We viewed this as a secondary impact measure. Although FLNP is not a behaviour management programme that teaches parents specific behavioural management techniques, it does aim to provide parents with a

wider and more skilful repertoire of parenting strategies both for managing challenging child behaviour and for reinforcing positive behaviours. Family Links were therefore keen to establish whether a robust measure of parenting behaviours could be found that reflected the programme's aims.

Parenting Young Children (PARYC; McEachern et al., 2012) is based on a measure developed for an American parenting programme known as the *Family Check Up*. It is comprised of 3 scales (supporting positive behaviour, setting limits; pro-active parenting) that map well onto FNLP content. Although the PARYC scale was originally designed and tested for children aged around 5-6 years old, we found it relatively straightforward to make it applicable to parenting children in a wider age range with only very minor modifications. We named the 16-item version *Parenting Our Children*, and it is shown in the Appendix to the full report along with the original PARYC for comparison.

(4) Parenting self-efficacy and coping

In practice, the terms 'self-efficacy' and 'coping with parenting' are used interchangeably in the literature. They are slightly different, but closely related. *Self-efficacy* for parents has been defined as the beliefs or judgements a parent holds of their capabilities to organise and execute a set of tasks related to parenting a child (De Montigny and Lacharite, 2005, p. 390). It concerns not only possession of skills, but also a person's beliefs that he or she can integrate them into an appropriate course of action. *Coping with parenting,* is defined by us as: *a parent's self-belief that they have the ability successfully to manage the tasks and other aspects associated with being a parent*. Both are strongly related to parenting stress and higher self-efficacy is known to underpin both parent and child wellbeing. Again, we viewed this as a secondary measure of overarching impact for the FLNP.

The extensive literature on this construct was reviewed in detail in a previous project to develop an OIM on precisely this area, and it was agreed to utilise the learning from this work. It was therefore agreed to use (without adaptation) the most successful measure that emerged from that work, the *Parent Coping Scale* or PCS (Ghate and Moran, 2013; based on Ghate and Hazel, 2002). This measure had already been demonstrated to have excellent psychometric properties and good sensitivity to change, and has been widely used in different variants in a number of countries in the world. Technical detail about the PCS is available on the Colebrooke Centre website[3]. It is shown there, and in the Appendix to the full report.

All of the measures, together with some demographic questions in Wave 1, were combined into a booklet for ease of completion.

[3] See http://www.cevi.org.uk/docs/Parent_Coping_Scale.pdf for technical information.

Psychometric testing of the four candidate measures

Alongside the main pilot in 30 Nurturing Programme groups, we tested the four finalised measures for their measurement properties and quality in a 'test-retest' procedure, to examine external reliability (stability) and internal reliability (consistency). Five groups (3 in the South, 2 in the North) agreed to ask a cohort of parents to complete the four measures on two consecutive weeks, and 27 parents provided usable data. We then analysed these data for **external reliability**, which refers to how reproducible the results are, or how stable the measure is, over multiple administrations. The test statistic used was a Pearson intra-class Correlation. Values of 0.70 and above are considered high or very high and indicate strong relationships between the successive administrations. We also tested for **internal reliability**, being a test of whether multiple items that make up a scale are internally consistent and measuring a single consistent idea. The test used here is Cronbach's Alpha. Values are typically considered good in internal reliability tests if they exceed 0.80[4].

The results of our tests are shown in the table below. They indicated that two of the measures had moderate external reliability and two high; and all had good internal reliability.

Table 1 Reliability test data for the four measures: 'Test-Retest' exercise

Measure	**External reliability/stability** (Intra-class correlations co-efficients between 1st and 2nd administration)	**Internal reliability/consistency** (Cronbach's Alpha; Test-retest exercise, data from 2nd administration)
1. *How Things Are in My Family*	0.61	0.86
2. *Family Climate Scale*	0.73	0.84
3. *Parenting our Children*	0.61	0.93
4. *Parent Coping Scale*	0.70	Not applicable (single item scale; assumed high)

The **concurrent criterion validity** of the individual scales can also be assessed by exploring the extent to which they correlate one with the other. This helps to assess how well an individual scale performs compares with another similar scale measuring similar phenomena. Although the scales were measuring different dimensions of parenting and family life, all were concerned with parenting, and to that extent, we would expect them to show associations with one another. To assess this we used the main pilot data.

[4] Bryman D and Cramer C (1999) *Quantitative data analysis, a guide for social scientists* London: Routledge

Correlation co-efficients were calculated to assess the degree of association between the various scales, and were carried out for each wave separately, as shown in the tables below. All of the correlations were significant at the p<.01 level. This suggests that all the scales were associated with one another to a degree, and as expected, *How Things Are In My Family* and *the Family Climate Scale* are particularly closely associated, whilst *Parenting Our Children* and the *Parent Coping Scale* are less well correlated (though still significantly so).

Table 2 Correlations between scales

Wave 1

	Family Climate Scale	*Parenting Our Children*	*Parent Coping Scale*
How Things Are In My Family	.82	.49	.44
Family Climate Scale		.62	.47
Parenting our Children			.42

Wave 2

	Family Climate Scale	*Parenting Our Children*	*Parent Coping Scale*
How Things Are In My Family	.73	.42	.34
Family Climate Scale		.60	.35
Parenting our Children			.26

Base = 131

Sampling for the pilot

The sample for the pilot was selected together with Family Links to provide the best possible test for the measures. Our aim was to test the measures as naturalistically as possible in all five regions/countries where the programme is run, and in a good cross-section of urban, suburban and rural areas. A list was therefore constructed by Family Links especially for the pilot, and the list of 65 groups thus formed was then categorised by region and urbanity.

Anticipating some attrition in the sample over successive waves, we purposively selected 35 groups in 22 different local areas, and in the event 30 groups took part, distributed as shown in the table 5. At Wave 1, these groups had a range of 4 to 15 participants in the pilot, with a mean average number of 9 participants returning data.

Method of administration for the pilot

For **Waves 1 and 2** (the pre- and post-test Waves) PGLs in the participating groups distributed booklets to all parents attending, and supervised their completion and return. At **Wave 3**, data were collected by the research team at Family Links and at the Colebrooke Centre using methods reflecting the personal preferences indicated by parents for re-

contact at the end of Wave 2. Some parents provided email addresses, some gave telephone numbers and some postal addresses, and were contacted accordingly, in order of email - telephone - post where parents provided more than one option. Email contacts were sent an invitation to participate in an on-line survey, and two reminders. Telephone contacts were telephoned by a member of the team and where possible, arrangements made for a fifteen minute structured telephone interview at a convenient time. Up to four call backs were made for this purpose. Those who provided only postal addresses were sent printed booklets for return by SAE by post, with no reminders.

Response rates

268 parents agreed completed a Wave 1 booklet, and 164 of these people subsequently completed a Wave 2 booklet; a very creditable **61%** response rate at Wave 2. Of these, **136** (51% of the Wave 1 participants) had provided enough **data at both Waves** for us to be able to send 'matched pairs' for data entry and analysis, and had given permission for us to re-contact them at Wave 3 (the rest had too much missing data to be suitable for analysis). The researchers attempted contact with all of these people by the variety of methods outlined above, and 50 completed Wave 3 returns were obtained. Thus the agreement rate to re-contact a third time given at Wave 2 was 83% (showing considerable goodwill towards the programme and the research) but in the event, the achieved response rate by Wave 3 was **37%** of the achieved W2 sample and 19% of the achieved Wave 1 sample.

Feedback from Parent Group Leaders

To explore the process of administration we asked Parent Group Leaders to provide feedback after each Wave. Seven PGLs provided feedback on pro-formas and seven by telephone (ie, just under half the total number of PGLS who participated).

There was strong appreciation of the purpose of the work, and the vital importance of collecting better and more defensible (i.e. representative) data on the effectiveness of the programme. The idea of an OIM, which would be short and easy to administer, was widely supported.

Most PGLs thought that parents broadly understood the purpose of evaluation, and understood that public programmes are funded on the expectation that they achieve positive change and 'help' those who use them. PGLs reported that very few parents seemed at all unwilling or apprehensive about completing the booklets, although some found reading difficult, and either asked for help or took a long time to complete the questions. Some of the longer scales looked daunting, and elicited exclamations from parents, especially at Wave 1 when they were seeing them for the first time. However PGLs noted that by Week 10, parents were "*much more relaxed*" about the process. (They also noted however that completion of the booklets sometimes took longer at Week 10, which

PGLs attributed to a change in levels of parents' reflectiveness by the end of the programme).

PGLs we spoke to reported they had carefully observed the procedures designed to keep parents' responses confidential (which included asking parents to seal booklets in envelopes before handing back) and thought that parents had understood that that the results would not be scrutinised by PGLs themselves. Nevertheless, some PGLs (a small number) reported suspicions of 'faking good' – parents wanting to please and thus reporting more favourably than was actually the case. However, PGLs who felt this was the case also noted that by Week 10, parents were more confident and more open. They worried that Wave 2 score might therefore be lower than at Wave 1, not because things had got worse, but because parents were able to be more honest and self-critical[5].

Results

Characteristics of parents and their children

Feedback from participating PGLs indicated that groups had, in general been fairly typical of the usual groups with which they worked. The tables below show the characteristics of the parents who took part and their families, based on the data provided by 136 parents who completed both waves of data collection at Wave 1 and Wave 2. The characteristics of the sample at Wave 3 (n43) is not shown here, but confirmatory analysis showed no substantial differences were detected in the demographic profile of this group as compared with the characteristics of the larger baseline group.

At baseline, the sample was predominantly comprised of mothers (89%), predominantly of White British ethnic group (84%). Approximately half of all parents were in the age range 25-34 years, and just under two thirds (62%) were living with a partner. The average number of children in the home was 2, and two fifths of families (44%) had pre-school children at home. Notably, three quarters of the parents (74%) said they had been recommended or referred to the Programme by a worker (e.g in a children's centre) and less than a fifth (17%) considered they had found out about it and referred themselves.

We asked parents to identify one child they would particularly like the Programme to help them with, to be designated as the focus of later questions that required answers about an individual child. We call these 'index' children. Index children were more likely to be boys (58%), were equally spread across age groups up to ten years old, and just under a fifth (18%) were reported by parents to have special needs of some type.

[5] A phenomenon noted in the survey methods literature as 'response shift bias' (see Moran and Ghate, 2013).

Table 3 Sex, and age and relationship status of parents (Wave 1 sample)		
	%	n
Sex of parent		
Female	89	121
Male	11	15
Age of parent (n134, 2 missing)		
Range	19-48 years	
Mean	31 years	
Under 25 years	19	25
25- 34 years	55	74
35 years or older	26	35
Relationship status (n133, 3 missing)		
Living with a partner	62	83
Living without partner	38	50

Base = n136 unless otherwise stated

Table 4 Family size and age of children of parents (Wave 1 sample)		
	%	n
Age of children in the home (n127, 9 missing)		
0-2 years	24	74
3-4 years	20	61
5-11 years	42	128
12-16 years	12	35
17 years or older	2	6
Family size (number of children)		
Range	1-6 children	
Median	2 children	
Index child's sex (n127, 9 missing)		
Female	42	53
Male	58	74
Index child's age (n127, 9 missing)		
0-2 years	20	25
3-4 years	25	32
5-6 years	24	31
7-10 years	20	25
11-13 years	6	8
14 years or older	5	6
Index child's needs (n127, 9 missing)		
Special needs	18	23
No special needs	82	104

Base = n136 unless otherwise stated

Table 5 Ethnicity and referral source of parents (Wave 1 sample)		
	%	n
Ethnicity (n134, 2 missing)		
White British	84	113
Asian/Asian British	5	6
Black/Black British	4	5
Mixed White and Black	1	1
Other	7	9
How parents were referred to the Programme (multiple answers possible)		
Worker (e.g children's centre, health visitor)	74	101
Friend or family member	15	20
Self-referred	17	17

Base = n136 unless otherwise stated

Results of the four measures

Headline findings were that despite reduced sample sizes over time, **all four measures detected statistically significant improvements between Waves 1 and 2, with these improvements maintained in all cases at Wave 3, statistically significantly so in two of the measures.** Thus all four measures were demonstrably sensitive to change and showed that parents reported positive improvements after having participated in the programme, continuing for at least some weeks afterward. Full details of the analyses and all test statistics are given in the full report. The tables below give snapshots of the results.

(1) How Things Are In My Family

Table 6 ***How Things Are In My Family***

Comparison of scores at Wave 1 and Wave 2

Total sample n=94	**Mean**	**(s.d.)**	**Paired t-test**
Wave 1	40.9	(7.4)	Wave 1 vs wave 2 scores
Wave 2	43.8	(5.9)	t= -4.52(93) p< .001

Base = 94

Table 7 ***How Things Are In My Family***

Comparison of scores at Wave 1, 2 and 3

Total sample n=26	**Mean**	**(s.d.)**	**Repeated measures ANOVA**
Wave 1	42.0	(8.0)	Wave 1 vs 2 vs 3 scores
Wave 2	44.9	(6.3)	F (1.6, 40.2) = 3.35
Wave 3	43.7	(5.3)	p=.055 [not significant]

Base = 26

(2) The Family Links Family Climate Scale

Table 8 ***The Family Links Family Climate Scale***

Comparison of scores at wave 1 and wave 2

Total sample n=131	**Median**	**Range**	**Wilcoxon Signed Ranks Test**
Wave 1	37	(14 - 49)	Wave 1 vs Wave 2 scores
Wave 2	40	(13 - 49)	Z = -4.53 p< .001

Base = 131

Table 9 ***The Family Links Family Climate Scale*** **Comparison of scores at wave 1, 2 and 3**		
Total sample n=42	**Median Range**	**Friedman Test**
Wave 1	37.5 (16 - 48)	Wave 1 vs wave 2 vs wave 3 scores
Wave 2	41 (25 - 49)	$X^2 = 4.43$ (2)
Wave 3	41.5 (24 - 49)	p=.11 [not significant]

Base = 42

Parenting our Children

Table 10 ***Parenting our Children*** **Comparison of scores at Wave 1 and Wave 2**		
Total sample n=118	**Mean (s.d.)**	**Paired t-test**
Wave 1	79.0 (16.6)	Wave 1 vs wave 2 scores
Wave 2	91.0 (13.1)	t= -8.96 (117) p< .001

Base = 118

Table 11 ***Parenting our Children*** **Comparison of scores at Wave 1, 2 and 3**		
Total sample n=35	**Mean (s.d.)**	**Repeated measures ANOVA**
Wave 1	77.1 (17.9)	Wave 1 vs 2 vs 3 scores
Wave 2	92.4 (13.1)	F(1.46, 49.69) = 25.7
Wave 3	92.7 (12.7)	p<.001

Base = 35

(3) **Parent Coping Scale**

Table 12 ***Parent Coping Scale*** **Comparison of scores at Wave 1 and Wave 2**		
Total sample n=131	**Median** **Range**	**Wilcoxon Signed Ranks Test**
Wave 1 score	3 (1 - 5)	Wave 1 vs wave 2 scores
Wave 2 score	4 (2 - 5)	Z = -6.66 p< .001

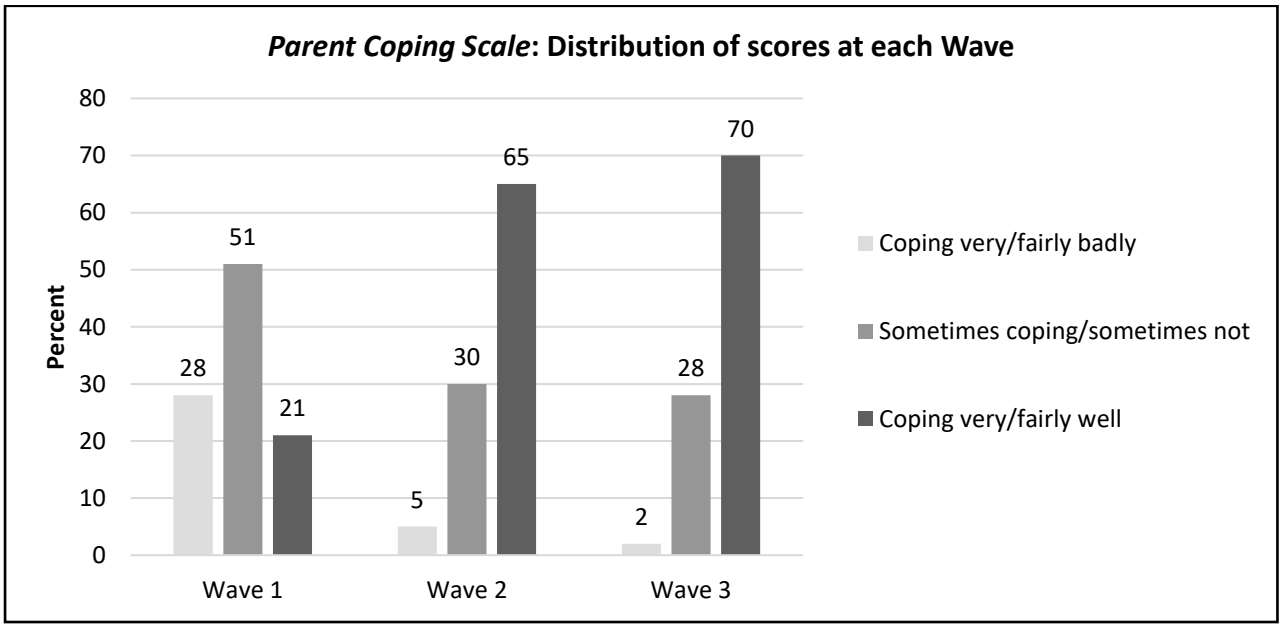
Base = 131

Table 13 ***Parent Coping Scale*** **Comparison of scores at wave 1, 2 and 3**		
Total sample n=43	**Median** **Range**	**Friedman Test**
Wave 1	3 (1 - 5)	Wave 1 vs wave 2 vs wave 3 scores
Wave 2	4 (2 - 5)	X^2 = 34.12 (2) p<.001
Wave 3	4 (2 - 5)	

Base = 43

A similar analysis for those who provided data at all three time points was conducted, using categorical rather than continuous data, and the figure below shows the improvement over time graphically:

Figure 8 Parent Coping Scale Changes in distribution of responses between Waves 1, 2 and 3



Base = 43

Some of the measures were easier to complete and more acceptable to parents than others: the shortest scale (the *Parent Coping Scale - PCS*) had highest acceptability and was completed by the greatest number of participants; and the longest scale (*Parenting our Children*) elicited most negative reaction and there were more missing cases and missing items for this question than for others.

Discussion

These pilot results were not intended to be used as evidence of the FLNP's wider effectiveness. However, the fact that all four picked up significant positive improvement over time, and maintained that improvement at the Wave 3 follow-up is extremely encouraging, not just methodologically speaking but also in respect of the prospects of demonstrating effectiveness of the programme in future evaluations 'at scale'.

We have no reason to doubt that the sample of parents who provided data for the pilot were representative of parents in general who complete the programme, and it is clear that for this sample, substantive positive improvement of statistically significant magnitude took place over the ten week programme, maintained up to twelve weeks later. These parents reported improved quality of family relationships and climate at home, more positive parenting practices, and improved perceptions of coping and self-efficacy. We took pains to maximise the confidential conditions for data collection, and we do not think that parents were 'faking good' to any substantial or lasting degree.

What we do not yet know, and what reviewers will point out, is whether that change would also have occurred for parents in a control or comparison sample (a 'counterfactual'); and whether other services that some may have been receiving might also have been contributing to the improvement. A next step would be to test the selected OIM(s) again, over at least a ten week period (corresponding to pre and post-test), incorporating a counterfactual that includes a 'no service' group, and perhaps also an 'alternative service' group to test how the programme performs comparatively.

Overall conclusions

The work to develop and test a group of potential **Overarching Impact Measures** for the Family Links Nurturing Programme formed part of a broader project of strategic research and development for the Programme. We successfully identified and piloted four measures of change in three domains of self-reported impact: family relationships/family climate; positive parenting practices; and parenting self-efficacy and sense of coping. 136 parents made up the pre- and post-test sample (51% of those who completed a questionnaire at Wave 1), and 50 of these parents completed a further questionnaire at Wave 3 (37% of the post-test group). All four measures were acceptable to parents (though to differing degrees), all were sensitive to change over time, and all showed significant positive change between pre and post-test, sustained at follow-up. Parents reported significantly improved quality of family relationships and climate at home, more positive parenting practices, and improved perceptions of coping and self-efficacy.

The measures, which were a combination of measures already pre-validated in previous research and adapted or new scales, were tested in 30 NP groups across five geographic regions/countries in the UK. Data were collected on the measures repeated at three time points or Waves: pre-test (before or right at the start of the first main session of the Programme; post-test (at the end of the final week of the Programme); and follow up (between ten to twelve weeks after the last session parents attended). A combination of methods were used, including paper and pencil completion during a Nurturing Programme group session, administered by PGLs; telephone administration by an interviewer; and self-completion by parents on line and by post. Psychometric properties were adequate-to-good for all scales.

The scale that worked least well in terms of administration and data quality (an adapted version of the FAD-12, which we called '*How Things Are in My Family*'), was not recommend for further use. However all three others can be used, subject to further testing and noting that two varied by demographic characteristics (as detailed in the full report). The *Parent Coping Scale*, though perhaps not compelling as the sole measure of impact, was the most robust and effective measure of the four, and is ready for immediate use. We recommend some further testing on *Parenting Our Children* (the adapted PARYC), and the *Family Links Family Climate Scale.* Norming and further 'cognitive testing' would be useful, and of course, the performance of all scales against a counterfactual (comparison or control groups) still remains to be tested.

The results were not intended to be used as evidence of the FLNP's wider effectiveness, but the fact that all four measures picked up significant positive improvement over time, and maintained that improvement at the Wave 3 follow-up, is extremely encouraging for the prospects of demonstrating effectiveness of the programme in future evaluations 'at scale'.

References

Bavolek S (2000) *The Nurturing Parenting Programs* OJJDP Juvenile Justice Bulletin Nov. 2000 US Dept of Justice: OJJDP

Epstein N., Baldwin L.M. and Bishop D.S (1983) 'The McMaster Family Assessment Device'. J. *Marital and Family Therapy 9:2* 171-180.

Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M. & Wallace, F. (2005). *Implementation Research: A Synthesis of the Literature.* Tampa, FL: University of South Florida, Louis de la Parte Florida Mental Health Institute, The National Implementation Research Network (FMHI Publication #231). http://nirn.fpg.unc.edu/resources/implementation-research-synthesis-literature

Fixsen, D, Blase K, Naoom S and Wallace F (2009) 'Core Implementation Components' *Research on Social work Practice 19* 531-540 http://rsw.sagepub.com/content/19/5/531

Gardner F , Hutchings J, Bywater T and Whitaker C (2010) 'Who benefits and how does it work? Moderators and Mediators of Outcome in an Effectiveness Trial of a Parenting Intervention' *J. Clinical Child and Adolescent Psychology 39 (4)* 568-580

Ghate D and Hazel N (2002) *Parenting in Poor Environments: stress, support and coping* London: Jessica Kingsley Publications

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

Hunt C with Mountford A (2003) *The Parenting Puzzle: how to get the best out of family life* Oxford: Family Links

McEachern, A. Dishion T.J., Weaver C.M. Shaw D.S. Wilson M.N and Gardner F. (2012); 'Parenting Young Children [PARYC]: validation of a self-report parenting measure'. *J Child Family Studies 21(3)* 498-511 doi: 10.1007/s10826-011-9530-y)

Moran P and Ghate D (2013) '*Development of a single overarching measure of impact for Home-Start: a feasibility study* London: Colebrooke Centre for Evidence and Implementation http://www.cevi.org.uk/docs/Impact_Measure_Report.pdf

Simkiss D E, Snooks H A, Stallard N, Kimani P K, Sewell B, Fitzsimmons D, Anthony R, Winstanley S, Wilson L, C J Phillips, Stewart-Brown S (2013) Effectiveness and cost-effectiveness of a universal parenting skills programme in deprived communities: a multicentre randomised controlled trial *BMJ Open 2013;3:*e002851. doi:10.1136/bmjopen-2013-002851

Stewart-Brown, S.L., Platt, S., Tennant, A., Maheswaran, H., Parkinson, J., Weich, S.,Tennant, R., Taggert, F. and Clarke, A. (2011) 'The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a valid and reliable tool for measuring mental well-being in diverse populations and projects'. *Journal of Epidemiology and Community Health,* Volume 65, Supplement 2.

About the Colebrooke Centre

The Colebrooke Centre for Evidence and Implementation (www.cevi.org.uk) is part of a new generation of intermediary organisations across the world, applying implementation science in real world practice improve services for children and families. The mission of the Centre is to improve the effectiveness of systems and services for children and families by promoting and applying an evidence-informed approach to their design and delivery. The Centre is founded on the recognition that high quality implementation is the key to better results, and that high quality implementation is evidence-informed. The Centre aims to harness the insights and tools generated in recent years by the movement towards evidence-based practice and implementation science for the benefit of the widest possible group of services and interventions.

The Centre is an independent not-for-profit company limited by guarantee (company registered in England and Wales, number 7712883). It is based in central London, and governed by a board of directors drawn from leaders in children's services, policy, research, and implementation practice and science. The Centre is funded by means of grants, contracts, consultancy and donations.

The Colebrooke Centre for Evidence and Implementation

55 St John Street
London EC1M 4AN
United Kingdom
www.cevi.org.uk

e: colebrookecentre@cevi.org.uk

t: +44 203 137 0486