

Development of an overarching measure of impact for Home-Start UK

A feasibility study

Patricia Moran and Deborah Ghate

Foreword

This report is welcomed as an important contribution to the debate surrounding how best to evaluate community based support for families in the 'real world'. While based on a case study of Home-Start, this report has significant implications for a wide range of service providers; for commissioners of local family support services and for policy makers deliberating on the future direction of family support in times of financial hardship. The authors rightly make clear this piece of work on the feasibility of devising a simple measure of overall impact for a service was seen from the outset as an adjunct to more comprehensive methods of independent evaluation. However, in the real world it is not always possible to undertake large scale evaluations but organisations still wish to have evidence that what they are doing is making a real difference to people's lives. This report contributes to the debate about how organisations can undertake self-evaluation in the most rigorous way possible given the resources available to them.

The learning from the process of undertaking this piece of work for both service providing organisations and those who would evaluate them, is that it is not simply a case of taking measures 'off the shelf', without careful consideration of what is being measured and for whom. Clearly understanding and articulating these aspects is a key step in developing the evaluation approach.

In particular, this feasibility study demonstrates how crucial it is to involve the insight and perspectives of the service provider to ensure that what is measured in realistic evaluation is both meaningful to and workable for the service. An example of this is the feedback from practitioners throughout this project which has been invaluable and has thrown light on the realities of data collection in peoples' homes and the implications for vulnerable families.

For wider stakeholders this report describes the methodological challenges of commissioning evaluations for universal access family support services where families' needs vary in intensity and range. It outlines the need for careful interpretation of evaluation results in the light of the appropriateness of the outcome measures chosen and it provides insight into developing low cost overarching service impact measurements. The project succeeds in demonstrating that with careful consideration of the constructs being measured, even in a universal access service, a promising reliable, valid and acceptable single impact measure can be developed. This is a useful approach for organizations to self-evaluate their work as part of an ongoing evaluation strategy.

For Home-Start this key report represents a further step in our commitment to provide excellent support for families in communities across the UK. Coping is a familiar and deceptively simple concept in everyday language. For families it can represent the difference between needing support and being independent. When a family is coping as a unit it can go on to achieve life changes for parents and children. Home-Start has built on these links between life changes for parents and children and the support Home-Start volunteers provide. We have developed a theory of change as part of our new strategic plan. The recent longitudinal evaluation of Home-Start Netherlands (Hermanns et al 2013) has also contributed to our thinking, giving as it does a clear exposition of the links between parental and child outcomes in the context of volunteer-led home based family support. Home-Start will continue to explore the usefulness of the Parenting Coping Scale which has been identified here as a promising measure for coping in families. This piece of work is an important foundation stone for this work alongside future external evaluations and interrogations of our large monitoring data set.

Wendy Rose OBE, Vice Chair of Trustees and Chair of Research Advisory Group, Home-Start UK
Rob Parkinson, Chief Executive, Home-Start UK
December 2013

Hermanns J., Asscher, J. Zijlstra, B. Hoffenaar, P. Deković, M. (2013) Long-term changes in parenting and child behaviour after the Home-Start family support program Children and Youth Services Review 35 (2013) 678–684.

Acknowledgements

This study was undertaken as a collaboration between Home-Start UK, Deborah Ghaté and the Centre for Effective Services. The study was designed, planned and implemented by Deborah Ghaté, Elizabeth Young, and Patricia Moran. Data were collected by Home-Start local schemes in Northern Ireland. It was analysed and written up by Deborah Ghaté and Patricia Moran.

The authors thank Wendy Rose, vice-Chair of Home-Start UK, and Dr Elizabeth Young, Director of Research and Policy at Home-Start, whose original ideas set us on the path that led to the publication of this study. Elizabeth and her colleagues Therese McCann (project manager) and Gillian Weston (data support) contributed to design and worked hard to facilitate the data collection, as did Heather Knox (director, Home-Start Northern Ireland) and her team of Co-coordinators who carried out the field testing. Elizabeth Young supervised the psychometric testing of the PCS during 2012. At CES we thank Dr Helga Sneddon.

Dr Karen Whittaker, Professor Tirril Harris, and Professor Larry Dumka all provided generous help at the review stage.

Deborah Ghaté, Patricia Moran
December 2013

About the authors

Dr Patricia Moran is an independent research consultant.

Dr Deborah Ghaté is the Chief Executive of the Colebrooke Centre for Evidence and Implementation. She is author of numerous research and evaluation studies on child and family studies, and works currently in the field of implementation science and practice. She was founding Chief Executive of the Centre for Effective Services between 2008 and 2011.

Executive summary

Voluntary and public-sector providers of family support are increasingly expected to provide quantifiable evidence of outcomes for service users at both national and local level. Yet the effort and costs of designing and carrying out evaluation studies to collect this information are substantial, and the results are often inconclusive.

This report describes the results of an innovative methodological development project to develop evaluation methods, conducted in collaboration with Home-Start UK, a substantial voluntary organisation providing befriending and support to vulnerable families, local Home-Starts in Northern Ireland, Deborah Gbate, and the Centre for Effective Services (CES). The aim of the project was to explore whether it was possible to develop a simple, overarching measure of the impact of Home-Start's work with vulnerable families, to use as a low cost evaluation tool. The project was conducted between 2010 and 2012, with fieldwork carried out in Northern Ireland during 2011.

This short summary of key findings is intended for practitioners and policy makers. A fuller technical summary for researchers and those interested in the methodological aspects of the project is available on the Colebrooke Centre website at:

http://www.cevi.org.uk/docs/Impact_Summary_2.pdf

Purpose of the project

The purpose of the project was to identify and test a short, quantifiable measure of impact that could be administered easily to capture the core goal or 'essence' of Home-Start services. The intention was to develop a 'big picture' measure that would apply to the widest possible proportion of the diverse group of families who use Home-Start. It was also intended that the measure we developed could be used by other comparable community-based family support services, and that the methodology we used could be replicated to produce alternative measures for other organisations with different goals.

Background to the project

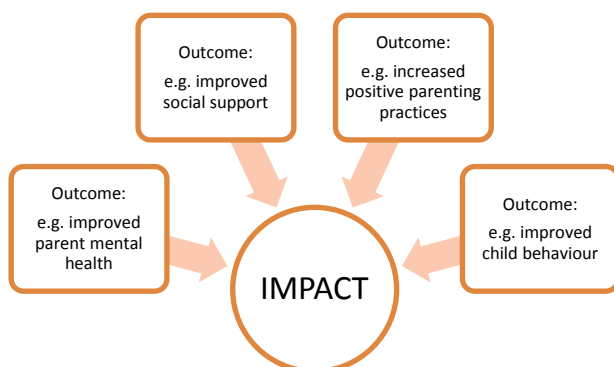
Home-Start UK is a substantial voluntary family support organisation. It has worked across the four nations of the UK since 1973 through self-governing local schemes (known as 'local Home-Starts'), providing volunteer support and befriending to families experiencing stress, who have children under five. In 2010-2011, local Home-Starts worked with around 17,000 volunteers and provided support to around 36,000 families with 77,000 young children. The dominant route by which families reach Home-Start is through referrals by health visiting, social work or other health and social care staff, although a substantial proportion of parents refer themselves. The service is based on universal access principles (ie, is not targeted on specific groups but open to all), and there are no fixed eligibility criteria that parents must satisfy in order to receive a service. The kinds of stresses experienced by parents vary widely, for example living with disabilities, living in poverty, or having inadequate social or family support.

Home-Start provides a service approach that is responsive, *user-led*, and tailored to individual parents' needs. Although the nature of the service provided to each family is underpinned by common principles, the precise help delivered by volunteers and the duration of this help is negotiated individually, and varies from one family to the next. This is an approach typical of many established family support providers across Europe, in contrast to more formal models of intervention that use programmed, *theory-led* delivery that aim to deliver a more standardised package of support to service users.

Because of the variability in the service provided to families, measuring outcomes using quantitative research methods is especially challenging for services like Home-Start. Quantitative indicators sometimes show little or no change, even when qualitative methods find that the service is extremely helpful to parents. This can be observed in some of the mixed evaluation findings for Home-Start itself over the years, and in many other instances in the wider family support evaluation literature.

Recognising these challenges, we made a distinction between *outcomes* and *impact*. Our aim was to identify a measure of *impact*, which we define as the overarching, high-level effect of a service that may be targeting a number of specific subsidiary outcomes. Impact can be thought of as the 'core change' that a service is trying to achieve, or the 'sum of the parts' of multiple outcomes. We were not attempting to measure *outcomes* themselves, which we define as changes over time in specific areas of learning, behaviour or life circumstances (e.g, parents' knowledge of child development, parents' methods of discipline, children's behaviour, mothers' mental health, increased social support).

Figure 1 A model for the relationship between outcomes and impact



Methods

The project had five key stages, which are described in more detail in the report and summarised in the box below:

Box 1 Methodology for the project

1. Literature review and consultation to identify the key aspect of the service to be measured (the overarching change that Home-Start hopes to achieve, which could best capture the overall impact of the service)
2. Literature review and consultation to develop the methodology, including exploring the advantages and disadvantages of simple as opposed to complex measurement approaches
3. Consultation and development of three alternative measures to be tested, involving adapting existing, widely used measures
4. Testing of measures by 12 local Home-Starts in Northern Ireland, with a sample of 76 parents, and three waves of data collection over a total average follow-up period of 23 weeks
5. Data analysis of results, and feedback from Home-Start co-ordinators on the process of administration, to explore the way the three measures worked
6. A final stage of further testing on one of the measures that proved most successful

Defining 'impact' for Home-Start

Desk research and consultation with key stakeholders was undertaken to identify the single most important aspect of parenting or family life that Home-Start was seeking to change through its help, which could be said to best capture the overall impact of the service. We were looking for an indicator of impact that was readily recognisable to the organisation's leadership, staff and volunteers as reflective of their intentions; applied to the widest possible group of parents who use the service; and closely reflected the direct, on-the-ground support that Home-Start volunteers give to families. To be practical for evaluation purposes, this 'change goal' should lend itself to simple quantitative measurement.

The review and consultations with stakeholders pinpointed the concept of parenting 'self-efficacy', and the plainer English construct of 'coping with being a parent' as most closely matching these criteria. Self-efficacy (sometimes called 'self-agency') in relation to parenting concerns the belief that a parent has of his or her ability to organise and carry out the tasks of parenting. Research shows that higher self-efficacy is associated with better quality of parenting. It is also closely related to coping, which is the ability to manage situations, tasks and the wider role of being a parent in the face of background stresses. Consultations with Home-Start stakeholders confirmed that enabling parents to manage stressful situations better, to feel more confident in their abilities as parents, and to take more enjoyment in parenting were key goals of the support offered.

Developing the measurement approach

A second review focused on the literature on research methods to identify quantitative measures with features that made them both practical to use, and scientifically robust. During this phase we also consulted with colleagues with expertise in measurement. We were looking for measures that were simple to understand and acceptable to Home-Start parents and volunteers (which generally meant, not strongly 'deficit' focused); brief and easy to administer at low-cost; suitable to be self-completed by parents; pre-tested preferably in larger or general population samples to give comparative data and the comfort of knowing the measures 'worked' in the field; and with good measurement properties, including *validity* (how well it measures what it sets out to measure) *reliability* (how reproducible or stable the findings are) and *responsiveness* (how well the measure detects change over time).

In addition, the relative advantages and disadvantages of single as opposed to multiple-item measures were assessed, since our goal was to identify the shortest, simplest and least burdensome measure possible. Although it remains a gold standard to use multiple measures in evaluation research, we found no compelling evidence against the use of well-constructed single item measures in appropriate contexts. The literature also indicated these may even have advantages, in certain circumstances, over longer or more complex measures.

Selecting and field testing the measures

On the basis of the two reviews, three measures were identified. After consultation with Home-Start Co-ordinators these were agreed for testing, with adaptations where necessary:

- A. *The Parenting Self-Agency Measure (PSAM)ⁱ* – a five item scale measuring parenting self-efficacy, developed in the USA but previously used in the UK, with a five point scale response format
- B. *Enjoyment of Parenting* – a single item measure on a five point response scale, widely used by various authors and a feature of many longer scales
- C. *Parent Coping Scale (PCS)ⁱⁱ*, adapted from a measure developed for a national study of parenting in Britain and widely used since – a single item ‘global coping scale’ that we modified into a five point scale

The three measures were combined into a short questionnaire, and twelve Local Home-Starts in Northern Ireland agreed to participate in field testing. Co-coordinators at each local Home-Start approached all parents starting Home-Start services from October 2010 to April 2011, seeking their consent to participate in the pilot study. Of 88 parents approached, 76 agreed, a response rate of 86% at Baseline, with characteristics representative of the wider population of Home-Start service users.

The short questionnaire was given by Co-ordinators to parents in the form of a self-completed booklet, which parents completed and then return to the Co-ordinator in a sealed envelope identified only by a serial number. This process was repeated at three time points, each approximately 10 weeks apart: Baseline (Time One): during an ‘initial assessment’ meeting; at a 10-week review meeting (Time Two); and at a 20-week review meeting (Time Three). In practice, the average total follow-up period was 23 weeks, and over time the numbers of participating parents reduced as families ceased to receive Home-Start services, or were unavailable for other reasons. This reduced the numbers to 51 at Time Two (67% of Baseline sample) and 34 at Time Three (45% of Baseline).

Results

All three measures were found to be acceptable to parents, and two of the measures (A and C) detected statistically significantⁱⁱⁱ change over time. *Enjoyment of Parenting* (Measure B) was found to be unsuitable for measuring change, due to a pronounced ‘ceiling effect’ (i.e. that proportions at the top of the scale at Baseline were already so high there was little room for improvement).

Measure A, the *Parenting Self Agency Measure*, showed modest change in a positive direction between all three time points, but the changes were statistically significant only between Baseline and Time Two. Measure C, the *Parent Coping Scale*, showed the greatest change in a positive direction between all three time points, with strongly significant differences detected by the final follow-up, despite the reduced number of parents providing responses.

Process feedback from Home-Start Co-ordinators

Although the Co-ordinators confirmed the overall validity of the constructs being measured, they also identified a degree of response bias ('faking good') by parents, especially at the Baseline stage. This was possibly due to parents being unwilling to disclose the real extent of difficulties at this initial stage of relationship-building with Home-Start, and being unclear about the purpose of the data being collected, given that Co-ordinators themselves were themselves distributing and collecting back the questionnaires. Inspection of the results tended to confirm this, especially for Measure A, which showed some evidence of a bias toward 'faking good' compared to data obtained using the same measure in other studies. Measure C showed Home-Start parents reporting more plausible levels of coping difficulties compared to a general population sample. However, response bias could also not be ruled out entirely for this measure.

Co-ordinators also confirmed that the methods of collecting the data must very simple and streamlined so as not to delay and interrupt the flow of work for co-ordinators, and to minimise the burden on participating parents. It was also apparent that for some, the process of administering the measures interfered unhelpfully with the complex process of establishing trusting relationships between Home-Start and the parents who use the services.

Conclusions and recommendations

This collaborative project was innovative in a number of ways, first in its attempt to develop a measure of high-level impact rather than focusing on domain-specific outcomes; second in the rigorous approach we used to identify, validate, develop and field-test alternative measures in collaboration with local Home-Starts; and thirdly in our use of feedback on the process from the staff who deliver the services.

The project has satisfactorily demonstrated that it is possible to develop and use a simple, low-cost overarching measure of impact for an open-access, user-led and community-based family support service delivered by volunteers across the UK. We were able to develop and implement a measure that was compatible with the service's broad intentions, acceptable to parents and staff, and satisfactorily robust as a measure of change over time. Two of the three measures tested proved able to detect change over a follow-up period; one, the *Parent Coping Scale*, showed greatest sensitivity to change and holds particular promise as a simple, low-cost measure that could be used by Home-Start and potentially also by other community-based services with similar high-level goals. The project also has much wider applicability, in having developed a workable methodology for the development of other measures, in the case of services that seek other, different kinds of overarching impact.

A limitation of the project is that the measures were not tested against a comparison group. This means that we do not know how the three measures might have performed in a group of parents who were not receiving Home-Start's services. It is possible that these measures would pick up the same degrees of change (or lack of it) irrespective of whether parents were receiving a service, or irrespective of whether that service was Home-Start or something else. It is also the case that a single overarching measure of impact is not a substitute for the more detailed and fine-grained evaluation of outcomes in specific

domains of parent and child functioning. However, where time and resources are limited, or where services need to assess their own high-level performance on an ongoing basis without access to external research support, this method may be an important complement to more comprehensive approaches to evaluation. Well-constructed measures of impact may also help towards the bridging of the disconnection that often emerges when qualitative and quantitative research are employed together to assess the results of family support services.

Specific recommendations

For future measurement of overarching, high level impact of Home-Start's services we can recommend the use of the *Parent Coping Scale* (PCS), which, of the three measures tested, appeared to achieve the best balance between practical, theoretical and scientific considerations. The PCS provides a global measure of 'coping with being a parent', and improving parents' sense of coping with parenting stress is recognised by Home-Start's stakeholders as a valid indicator of the broad intentions of the service. 'Coping' has a plain English meaning readily understandable to parents of all social and educational backgrounds. It is closely related to constructs such as self-efficacy and self-agency that have been shown in many research studies to be related to better and more confident parenting. Future development of the measure should ideally test its behaviour in a comparison (non-service) group.

Full details of the *Parent Coping Scale* (PCS), psychometric results from additional testing completed after the study, and how to use it, are available on the web at:
http://www.cevi.org.uk/docs/Parent_Coping_Scale.pdf

We also recommend that future use of the PCS should explore the use of telephone rather than face-to-face administration methods. This would remove the need for local Co-ordinators or other Home-Start personnel to administer the measure, reducing burden on the service and increasing confidentiality for parents. Co-ordinators would need to seek permission for ongoing telephone contact, but data could then be collected centrally, by an independent researcher or telephone research unit. This would allow for more accurate timing of follow-ups, and also allow collection of data from parents who have completely ceased to use the service. Longer follow-up periods could also be employed, increasing the size and usefulness of the data-set for monitoring the longer-term impact of the service. All of these factors, if implemented, may yield new insights into the workings of the PCS and will contribute to the further refinement of the methods described in this study.

i Dumka, L., Stoerzinger, H., Jackson, K., and Roosa, M. (1996) Examination of the cross-cultural and cross-language equivalence of the parenting self-agency measure. *Family Relations*, 45, 216-222.

ii Ghate D. and Hazel N. (2002) *Parenting in Poor Environments: Stress, Support and Coping* London: Jessica Kingsley Publishers

iii 'Statistically significant change' means change that is substantial enough to be unlikely to have occurred by chance.

For further information, see:

Full Report:

Development of an overarching measure of impact for Home-Start UK: a feasibility study

Moran P and Ghate D (2013)

available on the web at:

http://www.cevi.org.uk/docs/Impact_Measure_Report.pdf

http://www.home-start.org.uk/about_us/what_we_do/policy_practice_research

<http://www.effectiveservices.org/our-work/homestart>

Technical Summary for Researchers:

Development of an overarching measure of impact for Home-Start UK: a feasibility study - technical summary for researchers Ghate D and Moran P (2013)

available on the web at:

http://www.cevi.org.uk/docs/Impact_Summary_2.pdf

The *Parent Coping Scale*: background and technical information Ghate D and Moran P (2013)

available on the web at:

http://www.cevi.org.uk/docs/Parent_Coping_Scale.pdf

Home-Start UK www.home-start.org.uk

Colebrooke Centre for Evidence and Implementation www.cevi.org.uk

Centre for Effective Services www.effectiveservices.org

This project was undertaken as a collaboration between Home-Start UK, Deborah Ghate and the Centre for Effective Services

© 2013 Home-Start-UK, Deborah Ghate and the Centre for Effective Services

Table of Contents

	Page
Part one - background to the study	10
Section 1: Overview	11
Section 2: Home-Start: support and friendship for families	17
Section 3: Challenges in evaluating Home-Start and other user-led service models	20
Section 4: Previous research on Home-Start, and what it tells us	23
Part two - the study and its findings	28
Section 5: Identifying a construct to capture impact	29
Section 6: Selecting a measurement approach	35
Section 7: Selecting the measures to be tested	41
Section 8: Pilot testing of the selected measures	46
Section 9: Findings of the pilot study	49
Section 10: Conclusions, implications and recommendations	66
References	73
Appendices	80
Appendix 1: Parent questionnaire assessing outcomes	81
Appendix 2: Parental self-efficacy measures shortlisted but not selected	83
Appendix 3: Information sheet for parents	84
Appendix 4: Data collection process questions for Home-Start Co-ordinators	85

Part One – Background to the study

Section 1 An overview

1.1 Introduction

This report describes the background and conduct of a collaborative research project between Home-Start UK, a voluntary organisation providing befriending and parenting support through self-governing local schemes (known as 'local Home-Starts') across the four nations of the UK since 1973; Home-Start Northern Ireland; with part-funding from the Centre for Effective Services (CES), an independent not-for-profit centre for the promotion and support of evidence-informed policy and practice in child, family and community services in Ireland and Northern Ireland. Patricia Moran, independent research consultant based in London and Deborah Gbate, director of the Colebrooke Centre for Evidence and Implementation in the UK and formerly Chief Executive of CES, led the research design work, analysis and reporting.

The project was small-scale, but innovative, with potentially far-reaching applications. The aim was to explore the feasibility of developing and implementing an overarching measure of impact of Home-Start's work with vulnerable families. The objective was to identify a 'big picture' indicator that could capture the 'essence' of what Home-Start services aim to achieve for the parents with whom they work, and test this indicator in the field with local Home-Start teams to see if it worked, and how well. If successful, it was anticipated that such a measure could be used for self-evaluation by local Home-Starts, and as an adjunct to more comprehensive methods of independent evaluation.

When staff and Trustees at Home-Start first approached us with the idea for the project, the issues they hoped to address were predominantly strategic and practical. But there were also important future applications of the work, not just for Home-Start, but potentially for a much wider group of family support providers. In addition, the project presented the opportunity to contribute to understanding of a familiar, but little-theorised, scientific challenge in the evaluation of the impact of family support services. Below, we describe in more detail the background context for the project.

1.2 Developing evaluation methods for Home-Start: Strategic and practical issues

Home-Start UK has a longstanding commitment to making best use of research to assess and increase the impact of the organisation's work with vulnerable families. Also, like many other voluntary providers, in the context of growing emphasis on 'results-based accountability', they recognised increasing pressure from funders to provide quantifiable evidence of outcomes for service users at both national and local level. In addition, Home-Start had recently been developing their own internal systems for tracking and monitoring their work. As part of this, and notwithstanding their awareness of the methodological advantages of independent research and evaluation, they wanted to explore the possibility of incorporating methods for self-evaluation of effectiveness that could be controlled and administered by Home-Start itself. For all small-scale providers, such as local Home-Starts,

the costs and time burdens of working with independent research teams are substantial and can be prohibitive. Self-administered evaluation methods are therefore attractive, for reasons of both cost, and value for money. Not only are they less expensive, but they give greater control of the research to the service providers, making it more likely that the results will be owned and used by staff in daily practice. In the case of Home-Start, the fact that services are delivered mainly by unpaid volunteers only strengthens this case.

1.3 Developing evaluation methods for Home-Start: Scientific issues

In addition to assisting Home-Start UK to develop ways to use evidence to improve the effectiveness of its services, the project presented other important opportunities. First, the work offered the chance to develop a method with potentially wide applicability for the assessment of impact in a well-established service that embodies key features common to many family support services, some of which make the application of standard evaluation methods very challenging. These are connected to the provision of a flexible, tailored and highly individualised ‘service offer’ to families. Home-Start, like many other family support providers in both voluntary and public sectors, offers an *approach* to helping families, rather than a formal *model* of treatment. (See Section 2 and 3 for further details of how Home-Start works). Thus, it does not operate to a standardised programme which is determined in advance and administered according to a set structure, but aims to be highly responsive to the needs and wishes of individual service users. The support offered thus takes different forms with different families and, for this reason, its approach can be described as *user-led*. This is in contrast to a small but growing category of trademarked and syndicated ‘evidence-based programmes’ that are *theory-led*. These programmes are generally heavily influenced by the clinical sciences, are based on articulated ‘theories of change’, and are usually accompanied by a substantial body of documented procedures that aim to guarantee fidelity to an original design and to specific implementation quality standards (see for example, Utting et al, 2007). They generally focus on a few defined aspects of family or individual functioning (Mitchell 2011), even where they situate themselves within systemic or ecological frameworks. Home-Start, by contrast, considers as one of its defining characteristics that it adapts what it does according to the specific needs of each presenting family. Although there are common features to the approach, and common elements of the service provided, the work that is done is not offered as a professional service delivered by ‘experts’ following a pre-determined curriculum, but as personalised, responsive peer support with a strong focus on befriending¹ and doing whatever is practically useful to parents at the time. Most support lasts around 6 months and amounts to two to three hours per family per week, but this may be supplemented by additional support as needed, so that the time over which services will be offered, and the frequency or number of hours of contact time delivered may vary. Moreover, support is delivered by volunteers from the same local communities as service users. These volunteers do not necessarily have a professional background in a particular branch of social care, further increasing the variety of approaches brought to the support delivered. As a result, no two families receiving Home-Start receive an identical service offer.

¹ Home-Start’s leaflets and literature for service users describes the service as offering ‘Support and Friendship for Families’

The distinction between *theory-led* and *user-led* that we have developed here is an important one, and is informed by a definition of the term ‘theory’ following the helpful outline given by Whetten (1989), who notes that a properly developed ‘theory’ is composed of two essential ingredients - description and explanation. *Description* (which Whetten calls the ‘what’ in theory), refers to the identification of the specific factors relevant to the phenomenon being explored (e.g, for our specific purposes here, identification of the problem to be addressed, or the actions that could be taken to address it, or the outcomes for users that are targeted by the service in question), whilst *explanation* (the ‘why’) refers to the building of a logical model of how these factors stand in causal relationship to one another. The notion of causality is essential to the ‘explanation’ part of the equation: a credible theory or model must posit causal relationships allow the formulation of one or more testable hypotheses that can be empirically confirmed. This is achieved by explicitly delineating the relationships between the relevant factors and showing how they stand in logical relationship to one another. Thus we define a *theory-led service*, by this standard, as one that is structured and delivered firstly according to a particular identification of a need (or a problem) to be addressed by the service or intervention in question, and secondly, a verifiable explanation of how the actions taken as part of the intervention provided are causally related to meeting the identified need. In other words, a theory-led service has an explicit, articulated model (with “‘arrows’ to connect the ‘boxes’” as Whetten puts it) of how the service’s actions (inputs) are related to specific, pre-identified and measurable changes for service users (outcomes).

The design of Home-Start (and many other similar forms of family support provision) cannot be described as a ‘model’ in this formal sense. Home-Start is underpinned by careful observation over many years about what can help vulnerable families to manage their daily lives better. There is a distinctive ethos (of non-judgmental partnership with parents), a framework and a set of principles to which volunteers are trained and within which they must work, and explicit protocols for how volunteers must manage certain situations (such as concerns about child safeguarding, for example). However, these are not organised into an explicit ‘theory of change’ at the level of the individual service user. They might more properly be seen as constituting a set of service-level or organisational-level operating principles, which shape an approach to helping families that may take many diverse routes according to need, preference and the volunteer’s own skills and capacities².

In respect of its flexible and individually-tailored approach, Home-Start has much in common with many other providers of family support across Europe. However, this ‘real world’ flexibility creates considerable headaches for researchers when designing and analysing the results of evaluations. Evaluation science – especially the quantitative variety – values standardisation and homogeneity, and finds the messiness of the real world challenging. The identification of the appropriate outcomes to measure becomes challenging when *what* is being delivered cannot be defined, or what is being delivered differs from one service user to the next. Measurement schemes and their analysis suffer substantially when *how* the service is delivered varies from one service user to the next, for

² This is not to say that non-programmatic approaches to family support cannot be informed by theory (Social Pedagogy is a good example of one such approach), but simply to highlight the differences between services that operate according to a specific individual-level theory of change, and those that operate on a wider canvas.

example, in respect of hours of contact time, or frequency of contact. Both are made more challenging when service users are extremely diverse in terms of their presenting needs. Not surprisingly, when research is carried out into services that offer this degree of flexibility, the results are often hard to interpret and subject to vigorous debate.

1.4 *Impact versus Outcomes*

Scientifically, the project presented an opportunity to address a familiar but little-theorised methodological challenge in research on family support services: that is, the assessment of overall service *impact*, as opposed to the measurement of specific *outcomes* in different domains of functioning. Over the years, both in the UK and in other countries, Home-Start services have been the subject of a number of research and evaluation studies. Qualitative methodologies, using one-to-one in-depth interviews and group discussions, have consistently found evidence of positive overall impact on users and extremely positive perceptions of the services provided. However, studies using quantitative methods measuring change over time in specific outcome areas (for example, parental mental health, or child development) by means of standardised measures have typically been less positive, and have sometimes found minimal evidence of effectiveness. This disconnection between strong qualitative findings and weak quantitative results is a familiar one. There are many other examples in the family support literature of this phenomenon, where qualitative research shows that service users consider the service valuable and attribute to it important changes in their ability to manage difficult circumstances, yet companion pieces of quantitative research find little or no measurable change on key indicators: for example, in the UK, Barlow et al (2008); Family Links Research Team (2011); Biehal et al (2012). In the debates that follow, it is frequently impossible to determine whether the intervention is effective or ineffective, or which of the conflicting findings should be relied on. Practitioners (and indeed, service participants) often feel that the benefits of the work may have been missed or obscured by the research (Howard, 1980). Quantitative researchers themselves may also often sense, post-hoc, that the methods and measures they have used have in some way failed to capture the key dimensions of the service (e.g Stewart-Brown et al, 2011). Often, there is sense of not having been able to ‘see the wood for the trees’, as it were; in other words, that the attempt to zero-in quantitatively on highly specific outcomes may have created an overly -“narrow funnel” that obscures the wider picture of the overall impact of a complex service (Schorr, 2003).

The distinction made here between *outcomes* and *impact* is a key one. *Impact*, as we use it here, is a 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). In the case of this particular project, we aimed to develop a measure of ultimate impact: the ‘essence’, one might say, or the wider picture of what a service like Home-Start is trying to achieve. For this purpose we eventually determined upon the overarching construct of ‘coping with parenting’, as we describe in later parts of this report. We did not aim to measure any of the many short, medium and longer term domain-specific or person-specific outcomes of using Home-Start that might contribute to coping.

1.5 Summary of the approach to the project

The project began in 2010. It had five key stages, which are described in more detail in later Sections:

1. Literature review and development of the key construct to be measured
2. Literature review and development of the methodological approach
3. Development of the alternative measures to be tested
4. Piloting and testing of measures by local Home-Starts in Northern Ireland (three waves of data collection over approximately 26 weeks)
5. Data analysis and reporting

We began with two focused literature reviews. One was concerned with identifying appropriate 'constructs' for measurement; that is, the overarching aspects of parenting or family life that Home-Start might be seeking to change through its help, and that could best capture the overall impact of the service. This review covered evaluation literature on Home-Start itself, as well as the wider literature on parenting support and its outcomes. The second review focused on methodology, and specifically, the scientific and practical advantages and disadvantages of simplified and single measurement schemes. We also consulted with a small number of other expert researchers in the field at this point to seek external opinions on the strengths and possible pitfalls of our approach.

The results of the two reviews led us to select a number of alternative possible measures. These were developed into a short questionnaire, based on existing measures with some small modifications. These formed the focus of a consultation meeting with Home-Start Co-ordinators in Northern Ireland to gain their views on whether the constructs we had identified resonated with them as representing what Home-Start was trying to achieve in its work with families. After further changes reflecting feedback from Co-ordinators, field testing of two alternative measures began in 12 local Home-Starts, overseen by a project supervisor based in Northern Ireland.

Field testing was administered by Co-ordinators. All parents joining the service in these schemes from October 2010 were invited to participate, and where consent was obtained the questionnaire with its two alternative measures was administered on three occasions: at an initial visit at the commencement of service use, and twice more at 10 week intervals. Overall, 88 parents were approached to participate and 76 agreed (an 86% response rate at Baseline).

Data were collated by a member of the Home-Start UK's administrative team. When data were returned, these were analysed by the researchers, along with feedback on process aspects from Co-ordinators. The aim of analysis was not to 'evaluate' the impact of the service, but to assess the patterning and plausibility of the responses, and to explore strengths and weaknesses in the design of the measures and in the process of collecting data. Our conclusions in this respect are set out in Section 9.

1.6 Structure of the report

The report is divided into two parts. **Part One** describes the background to the feasibility study. It outlines the context for the study and some of the considerations that informed its design, including the important distinction that we make between outcomes and impact as described in Section 1 above. Section 2 of this part of the report describes Home-Start as a service: who uses it, and what is delivered. Section 3 reviews some of the methodological and practical/technical challenges associated with evaluating responsive, open-access services, and draws an important distinction between theory-led and user-led models of provision that is relevant for the rest of the discussion. Section 4 recaps on the existing research base on Home-Start, both qualitative and quantitative, and how this informed the search for a measure of overarching impact.

Part Two describes in detail the design and conduct of the feasibility study, its findings and its conclusions. Section 5 of this part of the report describes how an appropriate construct for measurement was identified, and teases out questions of impact ‘on what’ and ‘for whom’. Section 6 presents methodological and technical considerations in developing a measurement approach, and Section 7 describes the steps taken to select measures to be tested from amongst a number of existing alternatives. Section 8 describes the pilot testing of these selected measures in the field in local Home Starts in Northern Ireland, and Section 9 presents the findings. Finally, Section 10 sets out some conclusions and recommendations.

Section 2: Home-Start: support and friendship for families

2.1 Overview of Home-Start

Home-Start is a well-established family support service that provides early intervention volunteer support for families. It comprises Home-Start UK, which is an independently registered charity supporting 334 affiliated Local Home-Starts (or local schemes) across the UK. A sister organisation, Home-Start Worldwide, was formed in 1998 to offer similar assistance to families in other countries.

Home-Start offers free, confidential and non-judgemental support to families with young children who are experiencing stress. Support is provided by unpaid volunteers who are given training to visit families on a regular basis and continue to provide support. Typically, this is for up to six months, but it can be for much longer or shorter periods depending on the families' needs. The service aims to increase the confidence and independence of families by:

- Visiting families in their own homes to offer support, friendship and practical assistance
- Reassuring parents that their childcare problems are not unusual or unique
- Encouraging parents' strengths and emotional well-being for the ultimate benefit of their children
- 'Trying to get the fun back into family life' (Home-Start, 2010a).

Each local Home-Start is set up and run by a small local team, following a set of national guidelines. The local basis of schemes means that each service has familiarity with the local context and issues affecting families, as well as knowledge of the resources available in their community. The local Co-ordinators of schemes are paid, but volunteers who directly support families are unpaid. Figures for 2010-2011 show that local Home-Starts worked with around 17,000 volunteers and provided support to around 36,000 families with 77,000 young children. (Home-Start, 2012).

2.2 Families using Home-Start

Local Home-Starts generally work to locally agreed funding contracts (for example with the Local Authority). These contracts will often identify the 'type' of family which the service has to support, normally defined in a variation of Hardiker's tier of needs (Hardiker, Exton and Barker 1991) and predominantly at levels two and/or three (that is, families and children 'in need' and families and children with severe difficulties). Embedded within these contracts are often specific requirements for project work: for example supporting early learning, or tackling domestic violence or financial hardship. This work will be delivered within the Home-Start framework described in Section 2.4 below.

Parents can self-refer to the service or be referred by health, social or community workers. As many as a quarter of service users are self-referrals (Home-Start 2010a). The universal, open-access nature of the service means that it supports a wide range of families, from

those needing short-term support with a specific need, to those in complex family situations with multiple needs who may require longer-term assistance. A significant proportion are single parent families, living in rented housing and on low incomes (Frost, Johnson, Stein and Wallis, 2000; McAuley, Knapp, Beecham, McCurry and Slead, 2004). Home-Start aims to meet their needs in a way that respects the diversity of family structures and is culturally sensitive. The types of difficulties that parents ask for help with often involve coping with their own mental health problems; overcoming isolation and lack of support; managing their child's behaviour; coping with multiple births; or coping with their own or their child's illness or disability.

2.3 Volunteer support

Volunteers are usually parents themselves or else have a great deal of experience with children. They are recruited, trained and matched to a family by a local Co-ordinator. Potential volunteers attend a 10 week accredited introductory course that involves being interviewed in their own homes and screened. If selected, they receive further training and regular supervision from their Co-ordinator and on-going support from fellow volunteers. Once assigned to a family, volunteers typically visit on a weekly basis, for two to three hours at a time, for as many weeks or months as there are unresolved needs. Parents may also be offered the chance to attend parenting support groups along with other local families.

The support that volunteers offer is flexible and individually tailored, and the timing and content of visits are negotiated collaboratively with the family. Depending on the family's circumstances and needs, the volunteer's role may involve providing practical help with child care or around the house, as well as emotional support and advocacy with other local service providers. Home-Start's promotional material describes volunteers as providing:

- Precious time for listening and talking
- Help with the children
- A break for parents
- Practical help and reassurance
- A chance to meet other parents in similar situations
- Support to use local services and resources (Home-Start, 2010a).

The range of activities that volunteers and families engage in varies enormously, involving going shopping, going on outings, playing with the children, helping around the house, accompanying a parent to a medical appointment, or simply listening and talking.

2.4 Characteristics of the service

As noted in Section 1, Home-Start does not operate as a standardised programme of support. What is delivered in terms of emotional and/or practical support varies according to the needs of the individual family. It does however operate within a clear structure that includes accredited training for volunteers and staff, and a robust quality assurance system which has 8 standard quality criteria – see http://www.home-start.org.uk/about/quality_standards.

The description of the service provided above conveys some of the key characteristics of Home-Start's service. These include:

- An open-access policy
- Use of volunteers rather than professional practitioners as service providers
- Home-visiting (predominantly), rather than centre-based support
- Flexibility in relation to the type of help offered to individual families and a considerable degree of 'personalisation' in the service offer
- Emphasis on a non-judgemental relationship between volunteer and family
- Support for the whole family rather than either a parent or a particular child
- Open-ended support commitment.

The support offered by Home-Start may perhaps best be described as 'parent-led'. It is not based on any particular theoretical model of how change is achieved, but is determined by each individual parent and their own perception of their needs. By giving parents the opportunity to decide on the support they want and the way it is provided, the service aims to empower families. Characteristics of the service that have been identified in research about Home-Start as particularly valued by parents include: having someone to talk about matters that they would not want to discuss with professionals in statutory agencies; the non-judgemental nature of the support; and the confidential and trusting relationship that can build between family and volunteer (Frost et al, 2000; Kirkaldy and Crispin, 1999; McAuley et al 2004; MacPherson, Barnes, Nichols and Dixon, 2010). These aspects of Home-Start mean that it is particularly well positioned to reach the substantial proportion (Ghate and Hazel, 2002) of families in poor communities who are reluctant to engage with formal support services due to previous negative experience with professional service providers, or else fears about statutory intervention (Shinman, 2005). Research shows that these key characteristics of Home-Start are also recognised by professional referrers to the service. They particularly value its voluntary and non-stigmatising nature; the way that it complements statutory services; its flexibility; and its provision of a service where one would not otherwise be provided for families (Frost et al, 2000).

Section 3. Challenges in evaluating Home-Start and other user-led service models

3.1 Introduction

In the contemporary climate, evaluation of social care services has both scientific and political dimensions. The evaluation of effectiveness plays an important role in furthering knowledge of 'what works' within the field of parenting and family support. Increasingly, services that are funded from the public or private purse are also required to prove their value, and evaluation results play a pivotal role in building and sustaining the case for funding.

While Home-Start is highly valued by both service users and referrers, some of the characteristics for which it is most valued render it extremely challenging to evaluate. Indeed, as we set out in Section 4 below, numerous prior evaluations and research studies of Home-Start exemplify the problems outlined in Section 1 – most notably the 'disconnect' between the results from qualitative and quantitative enquires, and the raising of debate about the appropriateness of specific outcome measures used and whether they in fact had, captured what Home-Start was actually delivering 'on the ground'.

Home-Start is not alone in this: very many services come up against the same challenges when planning, executing and digesting the results of evaluations of their work. Therefore much of the discussion that follows, although based on the specific example of Home-Start, applies to family and parent support services across the UK and Ireland and beyond.

Below we highlight some of the main challenges for evaluation of a parent-led service such as Home-Start. It was an awareness of these issues that underpinned our decision to explore the feasibility of developing a single overarching measure of impact.

3.2 User-led versus theory-led service provision

As described in Section 1, a new generation of family support programmes that have evolved as part of the 'evidence-based movement' (Fixsen et al 2009) and have risen to prominence in the United States. These evidence-based programmes (EBPs) are theory-led, rather than user-led. They are based on an explicit 'theory of change', usually articulated through a 'logic model' or other explanatory documentation³. They are often described as 'manualised' because practitioners delivering the programme work from a manual or handbook, in which the specific content of each session, the number and frequency of sessions, and other aspects of how the work is to be delivered are set out. Practitioners are often highly qualified and are formally accredited to deliver the programme. They are supervised closely to monitor their adherence to the protocol. The programmes have generally been refined over time with reference to the results of a substantial body of research including numerous randomised control trials, and this is no accident: the programmes are *designed* to be evaluated, and any features that might make evaluation

³ For a glossary of these technical terms, see *The What Works Process: Evidence-informed improvement for child and family services* (The Centre for Effective Services 2011, Dublin: Centre for Effective Services)

difficult or impossible are purposefully modified. These programmes are therefore (in principle if not always in practice) reasonably straightforward to evaluate.

By contrast, although the service offered by Home-Start is underpinned by a clear set of core principles that are identified in background literature about the service (Home Start 2010a) the specific content of what Home-Start *is* and what it *does* is hard to pin down. Home-Start has evolved from a bottom-up approach to provision that involves listening to parents' concerns and responding to them on an individual family basis. Indeed, the original founders did not think in the modern terminology and paradigms of 'intervention models' but in terms of a practical, common-sense response to an observed need (Harrison, 2003). Home-Start's work with families - like many other family support services - is not based on an explicit model of practice underpinned by a specific theory of change, although there are a number of theoretical perspectives that potentially contribute to an understanding of its approach. Volunteers are trained and supervised, but the nature of what they provide is negotiated with individual parents rather than being shaped *a priori* by a theoretical perspective on what families may need to achieve change. In this sense, the emphasis is on a user-led definition of provision, and Home-Start is a 'service approach', rather than a 'programme'.

Services like this are notoriously challenging to evaluate (Moran, Ghate and van der Merwe, 2004; Ghate, 2001), and often, the first task of a research team is to construct, post-hoc and on behalf of the service being evaluated, a theoretical model that specifies the primary outcomes the service seeks to change, and core components or characteristics of delivery that are believed to lead to these outcomes (ie, the 'active ingredients') (Chen, 1989; Ghate 2001, Ghate et al 2008). Only when this is done can a sensible research design be developed and appropriate tests of effectiveness devised. The risk, however, of post-hoc theory construction is that evaluators, who may have little in-depth knowledge of how the programme is delivered on the ground, may misjudge the programme or impose their own expectations of how the programme should work. This may lead to mis-specification of the core components and the primary outcomes. Evaluation results, therefore, may be misleading and unrepresentative of what the service actually delivers.

3.3 Identifying appropriate outcomes for open access services

Another key characteristic of Home-Start that impacts upon its amenability to evaluation, and which is also the case for any universal service, is the variable threshold for access to the service. In addition to families with specific needs who may be identified in local funding agreements, any parent who feels stressed and has a young child can approach the service themselves or be referred for support. In practice this means that the types and severity of difficulties for which families request support can vary enormously. This contrasts with other forms of family support or intervention where, more typically, there are narrow criteria and higher thresholds for accessing support. Evidence Based Programmes [EBPs], for example, always target particular types of participants with particular types of difficulties in order to achieve a pre-specified outcome, such as a change in a particular aspect of child behaviour or parenting style. Eligibility criteria are carefully controlled and referral and acceptance protocols are rigorously enforced. Outcomes in these services are more easily defined since they relate directly to the specific criteria for accessing the service, such as severity of child behavioural difficulties. In the case of Home-

Start, individual families' circumstances vary so much that identification of a single set of primary outcomes that apply across all families accessing the service is more challenging. Statistically, in order for an evaluation to show that a service is effective, a significant proportion of service users need to have made an improvement on a specific outcome measure. If the outcome being measured is not equally relevant to all or most service users, then the overall results will be diluted. The service may appear to be ineffective, whereas in fact, the actual case may be that the service *is* effective, but only for some families under some circumstances. Unless the researcher has very large numbers of service users in the sample, to allow for fine-grained analysis of specific sub-groups, failure to select an appropriate assessment measure that taps into an anticipated change that applies to (or is salient for) all or most of the families using the service will result in overlooking positive results to some degree.

3.4 Honouring the 'non-judgmental' ethos of partnership with families

A key component of Home-Start's approach to working with families is its non-judgmental, collaborative and empowering ethos, which is at the heart of the service's 'parent-to-parent' support. Home-Start Co-ordinators are acutely aware that parents in stressful situations may respond badly to the administration of long, detailed questionnaires, interviews or observation procedures. Parents may not welcome visits from unknown researchers in addition to trusted service staff and volunteers. This potential for evaluation to 'interfere' with the service it assesses and undermine the trusting, confiding relationship that workers or volunteers strive to build with parents is an acknowledged risk in the evaluation literature (Zeedyk, Werritty and Riach, 2002). Although there are ways around these difficulties, the introduction of robust measurement as part of an evaluation of a service such as this needs careful consideration in order to honour the service approach and avoid disrupting critical alliances while still providing evaluators with the information they need.

The characteristics of Home-Start, as described above, give some indication of the difficulties that are encountered when the world of evaluation meets the real world of practice. Home-Start's particular brand of family support clearly does not lend itself easily to evaluation. Our decision to explore the development of a single overarching measure of impact does not, of course, resolve these difficulties: rather, it circumvents them, and other pitfalls may be introduced, which we discuss in more detail in Part Two. However, by focusing on broad impact, we are able to avoid getting ensnared in the difficulties of identifying specific outcome measures in favour of a 'big picture' measure that can be agreed by all – researchers and those who manage and deliver Home-Start – as representing the fundamental core aim of Home-Start services. By developing a simple, single measure that can be administered directly by those who provide the service, the introduction of potentially threatening methods of questioning are avoided.

Section 4 Previous research on Home-Start, and what it tells us

4.1 Introduction

In this section we review some of the most significant previous studies of Home-Start. We show how the studies and the debates to which they give rise exemplify some of the challenges outlined in earlier sections, and summarise what we can learn from them. We also set out the implications of their findings for the attempt to develop a single overarching measure of impact.

4.2 Qualitative investigations of Home-Start

There have been many small-scale, qualitative evaluations of Home-Start that focus on a local projects or a small number of projects in a geographical area, using a variety of approaches. Here we draw on those that implemented systematic approaches to the recruitment of participants, methods of data collection and data analysis. Qualitative interviews and focus groups with families, volunteers and referrers are particularly useful for providing detailed description of families and their experience of the service, as well as exploring the nature of outcomes.

Descriptions of the needs of families are provided by several qualitative studies (Frost et al, 2000; McAuley et al, 2004; MacPherson et al, 2010). They indicate that most families accessing Home-Start are socially and economically vulnerable, and are dealing with multiple, competing demands. Typical stresses include parental isolation, maternal mental health, financial strain, parenting issues, children's special needs and maternal health/disability issues. Some are long-term, chronic difficulties and others are shorter-term crises. However, while these stresses are common among families accessing Home-Start, qualitative studies nevertheless indicate a great degree of variability in families' contexts and the background to their difficulties.

Studies following up families who have received Home-Start's support over time indicate that parents report improved well-being, enhanced self-esteem and confidence, improved relationships with family and friends, and reduced parenting difficulties (Frost et al, 2000; McAuley et al, 2004). McAuley and colleagues found that among families receiving Home-Start's support, most parents firmly attributed a reduction in stress to the support they had received from their Home-Start volunteer.

4.3 Quantitative investigations of Home-Start

Like qualitative studies of Home-Start, quantitative studies also vary in scale and robustness of design. Below we focus on recent studies that have involved pre- and post-intervention ('before and after') assessments of families receiving Home-Start and that have contrasted results with a group of families who have not received the service. This type of research design allows for greater confidence in attributing changes among families to the effects of the service.

In a study in 2004, McAuley and colleagues compared families receiving Home-Start over an eleven month period with a comparison group of similar families not receiving the service (McAuley et al, 2004). The choice of outcomes for investigation were derived from a preliminary study that took into account patterns of referrals to Home-Start, parents' views, the views of child welfare academics, as well as findings from relevant literature (Family Support Outcome Study, McAuley, 1999). This led to assessment of Home-Start's effectiveness in relation to the following outcomes: maternal mental health, parental stress, maternal self-esteem, maternal social support, and child development. In spite of the positive qualitative findings, quantitative results showed that there was no greater benefit for families receiving Home-Start support, because improvements on several outcome measures were also reported among comparison families.

In 2006, Barnes, MacPherson and Senior used a cluster randomised design to allocate families to study groups based on whether they were receiving Home-Start services or not. Women were screened in late pregnancy and selected for the study on the basis of a social disadvantage index score (a departure from the usual method for families to access Home-Start). A three-way comparison was drawn between Home-Start families receiving at least two visits, those who were offered the service but received one or no Home-Start visit, and a control group who were not offered the service. The outcomes assessed included maternal depression, parenting stress, mothers' perception of social support, the home care environment, health service uptake, and infant health, behaviour and development. Assessments took place when babies were two months and twelve months of age. Findings indicated that, overall, there was limited evidence of service impact. Two differences emerged that were in an unexpected direction: supported mothers reported giving fewer healthy foods to their infants at 12 months; and infants in families receiving Home-Start support have lower cognitive development than those in the control group. However, some significant positive effects of Home-Start support were also identified: supported mothers reported a greater proportionate decrease in parent-child relationship difficulties compared to control mothers, though the absolute level did not differentiate the groups at either time point; and mothers who received more visits and were supported for longer reported greater informal support from friends and family at 12 months.

A more recent study that took place in Holland also involved comparison of recipients of Home-Start with a control group, and is reported in a series of papers (Asscher, Deković, Prinzie, and Hermanns, 2008a; Asscher, Hermanns and Deković, 2008b; Asscher, Hermanns, Deković and Reitz, 2007; Deković, Asscher, Hermanns, Reitz, Prinzie and van den Akker, 2010). Families were assessed at four points: prior to receiving the service, one month after starting to receive the service, six months after receiving the service (i.e. at the immediate end of receiving the service), and at follow-up one year after the start of receiving the service. The outcomes assessed included: maternal depression, parental competence (or efficacy), parenting behaviour and child behaviour (including self-report and observational assessments of the latter two constructs). The findings indicate that as a group, Home-Start recipients improved in maternal efficacy (feeling competent to be a mother), and also in relation to two aspects of parenting behaviour, consistency and sensitivity (responsiveness to the baby's cues). There were no significant changes in other aspects of positive parenting behaviours, or in any negative parenting behaviours (Asscher et al, 2007). Structural equation modelling demonstrated that take-up of the Home-Start service led to an increase in maternal sense of competence that contributed to an increase

in supportive parenting, but this was not associated with a change in outcomes for children (Deković et al, 2010). Further analysis, using a statistical approach exploring change for individuals rather than groups, indicated that Home-Start mothers made significant improvements in maternal competence and well-being in comparison to the control group. The authors found that there were two types of families that the service affected differentially: an initially more needy group who showed a good deal of improvement but who did not reach community level functioning at the end of receiving the service, and a second group who did recover to community level function but for whom the overall *degree of improvement* was less because they were functioning closer to the community level in the first place, before receiving the service (Asscher et al, 2008a).

4.4 What do the studies tell us?

It is clear that previous evaluations of Home-Start have produced mixed results. In quantitative studies, positive outcomes (both statistically significant and not significant) have been found in relation to: parental 'efficacy' (or competence); parent practices including consistency and sensitivity (Asscher et al, 2007; Barnes et al, 2006; McAuley et al, 2004); and social support (McAuley et al, 2004; Barnes et al, 2006). However, each study tended to find somewhat different patterns of positive results, so that there is not consistent evidence of replicated good results. There were also a fair number of null findings and some negative ones. Qualitative studies, as noted earlier, tended to find more consistently positive effects.

What can we make of these mixed results? In respect of the Dutch studies, where results were generally more consistently positive, the authors suggest that enhanced parental sense of competence could be an 'active ingredient' of Home-Start that is responsible for bringing about change among participants. They also suggest that the lack of impact on child behaviour outcomes may be the result of the need for a longer follow-up period to allow for increases in more supportive parenting to take effect. In a review of the evaluations by MacAuley and colleagues and Barnes and colleagues, Barrett (2007) suggests that certain aspects of these study designs may have contributed to the lack of evidence of impact. Both studies involved 'pre-intervention' assessments taking place *after* intervention had begun, and lacked longer-term follow-up to allow for the possibility of 'sleeper' effects. In relation to McAuley and colleagues' study, Barrett identifies limitations related to statistical treatment of families receiving the service as homogeneous when there were differences in families in geographic locations associated with referral practices, types of families and presenting problems. There were also difficulties with measures, such as use of a measure designed to assess child outcomes for only one child under age three years per family, when Home-Start aims to make a difference to all of the children in the family and does not target work on a specific child. Also, in relation to Barnes and colleagues' study, participant recruitment on the basis of a social disadvantage index score was a departure from the usual method of recruiting families to the service and thus was not representative of how Home-Start works in practice. It is also not known for all of the studies how certain characteristics of families who were offered the service but did not take it up might have differed from those who did (for example, in relation to self-help attitudes or ability to make changes at home), and these types of factors may well influence outcomes in a range of ways.

4.5 Implications for the identification of an overarching measure of impact for Home-Start

The limited evidence of effectiveness in previous studies raises the possibility that Home-Start may be ineffective, or only weakly effective. However, if this were the case, one would expect uniformly null or negative results from the research (which is *not* the case). The positive qualitative findings also tend to raise objections to this counsel of despair. Given the great challenges presented to evaluation design and implementation of highly responsive, non-standardised, user-led services, other possibilities must also be considered.

We would argue that the mixed results could, in part or in whole, be due to the difficulties of selecting specific outcome measures that adequately capture the impact of a service like Home-Start, which is parent-led rather than theory or model-led. It may also be possible that the introduction of formal evaluative procedures alters the delivery of the service in unproductive ways, especially when the service ethos is founded on non-judgemental, non-stigmatising, informal engagement with families on a peer-to-peer basis. Findings by the Dutch teams, who experimented with more fine-grained analyses, also tend to bear out our earlier comments that if a service has differential effects for particular sub-groups or types of family, then analysis of *average* scores for the *whole* intervention group will statistically dilute these results. (Treatment effects are more likely to be statistically significant when the whole intervention group changes by a similar amount in a similar direction). Unlike participants in an evidence-based programme who share common eligibility criteria and receive a more standardised model of service provision, Home-Start families are a very diverse group experiencing stress at different levels and in different forms. They are also receiving differing forms of assistance in response to their individually-expressed needs rather than in a carefully controlled 'treatment package'. As a result, there may be considerable variation in responses to the service, which makes detection of effects more difficult (see also Harris, 2009). Until these possibilities are ruled out, it would be premature to pass judgment on the service's effectiveness.

It will be clear from the foregoing discussion that in the course of deploying evaluation science to understand a service's effectiveness, almost endless arguments and counter-arguments can be marshalled and endless refinements to standard evaluation technologies can be demanded. Fine-grain approaches to measurement of constructs, taking into consideration the characteristics of the intervention and its participants, using comparison or control groups, assessed at multiple time points, and analysed using a range of statistical techniques may be required. But the cold facts are that such complex methods are costly and time-consuming, sometimes to a disproportionate extent in relation to the overall investment being made providing the service. In the case of many family support services, these approaches simply are not feasible, and may never be so.

Therefore, an alternative approach, and the one that we explored in this project, may be to focus on impact, rather than outcomes: that is, to find measures that relate to an overarching construct that has broad applicability to the greatest possible number of

families using the service. After review of the relevant literature, our strong impression was that this could account for the success of the construct of 'parental efficacy' or competence in demonstrating the impact of Home-Start in the evaluation carried out by Asscher and colleagues. The present report considers the choice of a construct to measure overarching impact for Home-Start in further depth, and also considers specific approaches to its measurement, given that measurement itself may negatively impact upon a service.

Part Two – The study and its findings

Section 5: Identifying a construct to capture impact

5.1 Introduction

In this part of the report, we set out the findings of the five key stages of the project. We begin in this Section with a discussion of the findings from the literature review that was undertaken to inform our choice of overarching measure of impact for Home-Start.

Our search for an appropriate construct that would capture the fundamental, overarching aim of Home-Start's work with vulnerable families was informed not only by previous studies of Home-Start itself, and discussions with Home-Start staff, but also by a short and focused review of the wider theoretical research literature on parenting support and its outcomes. Two key questions formed the starting point for this investigation of appropriate measures of overarching impact: (1) impact *for whom?* and (2), impact *on what?*. 'Impact for whom?' is a question about whether the service is primarily aiming to support children or parents, and therefore whether measurement of overarching impact should focus on parents (for example, their behaviour, or their wellbeing), or children (for example, children's health, development or wellbeing). 'Impact on what?' follows the answer to this question, and invites us to be more specific about the nature of the changes the service aims to bring about.

5.2 Impact for whom?

The question of whether the service seeks impact for parents, children (or both), turns out to be a complex one. It was clear from our discussions with staff and volunteers that although Home-Start volunteers work within the family, spending time with parents and children, their actions are targeted primarily at supporting parents to manage stressful situations. Implicit in this approach, however, is the expectation that better supported, happier and calmer parents will parent better, and that this will in turn lead to better outcomes for children: that is, in the language of theories of change, that parenting support acts as a 'mediator' of child outcomes (in statistical models, mediators are factors that account for the effects of an intervention, by specifying how or why the effects occur, and are to all intents and purposes the 'active ingredients' that contribute to an intervention's success).

With this in mind, we concluded that for the purposes of this project, an overarching measure of impact that properly reflected the fundamental purpose of Home-Start should focus on parents rather than on children.

5.3 Impact on what?

We wanted to ensure that any construct suggested as a potential measure of overall impact should be defensible as part of the implicit model of how Home-Start works to deliver change for families. To this end, to answer the question 'impact on what?' we began by re-examining the research on Home-Start to identify what outcomes had been found to show change in any of the studies. One outcome area that stood out clearly as a fruitful line of enquiry was 'parenting efficacy' – a concept that is also referred to by a variety of other

related terms such as self-efficacy, coping, 'managing' and 'adaptation'. Below, we define and describe the different variants of this construct, and outline how the literature led us gradually to the final selection of an appropriate construct.

5.3.1 Coping and stress

A broad theoretical area prominent in the literature on family support that has much relevance for the work of Home-Start is that of coping with stress, as all of the families that Home-Start works with are experiencing substantial stress in some form. There are a number of theories that connect stress with coping. They take either the individual or the family as their focus, largely depending on whether they stem from psychological or sociological literature. In the case of stress and coping in relation to the *individual*, stress has been variously defined as: the internal state of the person; the effect of an external event such as a major life event or everyday hassle; and as an experience resulting from an interaction between a person and the environment (Mason, 1975). The concept of coping has shifted over time from being viewed as a 'style' or relatively fixed personality trait for dealing with stressful situations (ie, one is either 'a coper' or one isn't), to being viewed as a process that changes over time and in relation to specific situational demands (i.e., different people cope more or less well at different times and depending on what is happening in their life) (Lazarus, 1993). Coping has been defined as a person's efforts through thoughts and behaviour to manage demands that they perceive as overwhelming or extremely taxing to their personal resources (Folkman and Lazarus, 1991). Viewed in this way, coping can take the form of *problem-focused coping*, aimed at dealing directly with the source of stress, or *emotion-focused coping*, aimed at addressing the feelings evoked by the source of stress.

An influential theory of how coping works put forward by Lazarus and Folkman (1984) suggests that an individual's response to a stressor such as a life event or crisis involves two thought processes: *primary and secondary appraisal*. Primary appraisal concerns the person's judgement of the stressor as either a threat (and potentially harmful) or as a challenge (with the potential for enabling growth). Secondary appraisal involves the person's judgement of whether he or she has the resources needed to deal effectively with the stressor. Another distinction in the coping literature is made between *coping resources* and *coping responses*. The former concerns what is available to people to use, regardless of whether they use it, and the latter concerns what people actually *do* in specific stressful situations. Different coping responses may be required in different circumstances on the basis that some forms of coping have been shown to be useful in improving outcomes in some situations, but detrimental to outcomes in other circumstances (Lazarus, 1993).

Studies of stress and coping regularly appear in the psychological literature (Altmaier, 1995), typically focusing on very specific sets of circumstances, such as coping with ill-health or bereavement. Despite much investigation of the construct, coping in the context of parenting lacks a general definition, and there are many unanswered questions regarding the way in which quality of parenting may be affected by different approaches to coping, or how outcomes for children may be affected by parental coping. The few studies of parental coping that have been carried out tend to relate to very specific difficulties such as children's illness or disability. The extent to which their findings can be generalised to parents facing other types of problems is unknown.

Much of the stress and coping literature treats the coping *process*, and the *outcomes* of coping, as conceptually distinct. Several hundred 'ways of coping' have now been identified and many questionnaires have been designed to assess them (Skinner, Edge, Altman and Sherwood, 2003). They are typically used in studies assessing the relationship between choice of coping strategy and outcomes such as mental health. While the separation of coping responses from coping outcomes is important in the context of investigating how different ways of coping with a specific situation influence different outcomes, it seemed to us that considering *coping as a result (outcome) in itself* would be most useful for our purposes. We were less interested in the detail of how parents coped with different problems, and more interested in whether their overall ability to manage parenting (to 'get through the day' in more every-day language) was affected by Home-Start's support. When the concept of 'coping' is used like this, it has much in common with two other important concepts in the parenting and child development literature: those of 'resilience' and adaptation (Ghate and Hazel, 2002).

Home-Start's stated aims include offering families support and friendship, encouraging parents' strengths and emotional well-being, and building families' links with wider support and service networks. When viewed in theoretical terms, these aims have much in common with strengths-based approaches, since social support (including informal and formal support) and parental strengths (including psychological resources such as self-efficacy) are coping resources that may enable family adaptation.

These constructs therefore represent a promising avenue for exploration as possible outcomes for Home-Start. *Self-efficacy* has been chosen as a particular parental strength because of its hypothesised links with social support, its theoretical significance, and as a construct that has been shown to change as a result of Home-Start's intervention.

5.3.2 Parental 'self-efficacy'

Self-efficacy has been identified as a component of a person's thinking that influences coping. Self-efficacy is a construct developed by Bandura (1977), and is concerned with a belief in one's abilities to successfully perform a particular behaviour. Self-efficacy, in Bandura's model, is specific to a particular task or situation (ie, is 'domain-specific'), and is connected with the belief that a person has of their own the ability to influence an outcome of a particular situation. Self-efficacy has been linked to the 'secondary appraisal' process, in which an individual considers whether or not they have the resources needed to deal effectively with a stressor.

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 is often used interchangeably in the literature with related concepts such as parental competence, confidence and self-esteem. However, more careful consideration of these related terms suggests that competence is a *pre-cursor* of self-efficacy, as competence concerns possession of skills, but efficacy 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 (De Montigny and Lacharite, 2005).

The idea of 'parental self-efficacy' is concerned with how parents assess their own capabilities, rather than how parents behave *per se*. There is evidence from research that beliefs about self-efficacy can be altered by parent support and training programmes (Pisterman, Firestone, McGrath, Goodman, Webster, Mallory and Goffin, 1992; Bloomfield and Kendall, 2007).

According to Bandura (1977), a person's self-efficacy beliefs develop in relation to four primary sources of information. These include: 'enactive mastery experiences' (such as one's past successes or failures), 'vicarious experiences' (obtained through observation and copying others), verbal persuasion (including feedback from respected others who believe in the parent's capacities), and physiological and emotional states (physical or emotional symptoms that occur as response to changes in the brain and body when under stress). It is possible that the support that parents receive from Home-Start's volunteers influences all or some of these self-efficacy information sources. For example, volunteer assistance might: enable parents to have more success in managing the family and hence help foster a sense of mastery; provide opportunities for parents to observe volunteers' modelling of parenting practices; provide verbal reassurances of capacity to cope; all of which may reduce the sense of feeling stressed.

Research directly related to Home-Start provides some evidence that improved parental self-efficacy could be one outcome for parents using Home-Start, in the study by Deković and colleagues, as discussed in Section 3.4. This study used a measure of parental self-efficacy and showed that enhanced self-efficacy led to an increase in consistent and responsive parenting (Deković et al, 2010). Other studies have suggested that parental efficacy is linked to greater maternal sensitivity (the ability to interpret the 'cues' given by infants to elicit care giving behaviours), warmth (Teti and Gelfand, 1991), responsiveness (Stifter and Bono, 1998), and less use of harsh discipline (Sanders and Woolley, 2005).

Deković and colleagues also found that the social support offered by Home-Start's volunteers boosted parents' self-efficacy and positively influenced parenting practices. Other studies also provide evidence of a mediating role for parenting self-efficacy in relation to social support, and parenting behaviours. For example, a US study of Mexican immigrant mothers found that social support was related to parenting behaviours in part because those with greater social support felt more effective as parents (Izzo, Weiss, Shanahan and Rodriguez-Brown, 2000). A prospective study of women assessed during pregnancy and three months later found that social support protected mothers against depression through the mediating effect of self-efficacy (Cutrona and Troutman, 1986).

5. 4 Conclusions: Implications for choice of an overarching measure of impact for Home-Start

Our search for a suitable construct to represent the overarching impact of Home-Start's services required that we identify a factor that would have wide applicability to all or most of the families with whom the service works. It also required that the construct be consonant with Home-Start's stated aims and the purpose of the service as understood by the people who deliver it. This work has involved consideration of findings from previous research carried out on Home-Start and a review of the wider theoretical and empirical literature of relevance to parenting support. Although the complex relationships between stress, social support, parental self-efficacy, parenting practices and child outcomes require further investigation by robust empirical research, our conclusion is that self-efficacy and coping are promising constructs for several reasons.

In theoretical terms, parental self-efficacy is described as a 'coping resource' that enables parents to manage and adapt to stressful circumstances. In this, it has clear relevance to Home-Start's stated aims of strengthening and enabling parents who are parenting under stress. It is a construct that has broad application across all families, regardless of family structure, age of children, or type of stressful circumstances that a family may be dealing with, which is appropriate for Home-Start given the range of family circumstances that volunteers encounter.

At an empirical level, we also found evidence that as a broad measure of impact, parenting self-efficacy is shown to be associated with some very specific and important results or outcomes of parenting support. Thus, parental self-efficacy has been shown to relate to parenting practices and the quality of parenting. Finally, we have evidence of relevance of the construct from actual studies of Home-Start as well as of other types of parenting support programmes and services. There are many areas for further investigation in relation to parental self-efficacy, including the need for more conclusive evidence on its relationship to child outcomes. Further exploration is also required of the way in which social support and other factors influence self-efficacy in parenting. However, given these caveats, parental self-efficacy appeared to be a promising construct for our purposes.

Having arrived at this conclusion based on our review of literature, the final stage of our investigations involved a consultation meeting with Home-Start Co-ordinators to test out the construct against their own understandings of what Home-Start aimed to do. Was the construct of self-efficacy relevant? If a parent's sense of self-efficacy improved as a result of

contact with Home-Start, did this capture the essence of what the service hoped to achieve?

Not surprisingly, the terminology of 'self-efficacy' was considered by Co-ordinators to be relatively off-putting and alien. However when rendered into the more familiar plain English terminology of 'coping with being a parent', Co-ordinators agreed that this was indeed what the service hoped to achieve. They might not have the power to manipulate the objective circumstances in which parents found themselves, but if as result of the support and friendship offered by Home-Start volunteers, parents felt better able to cope with stress, and with parenting under stress, a satisfactory service had been provided. Support for this position also came from qualitative findings in one of the previous studies of Home-Start. In MacAuley and colleagues' report of follow-up interviews with parents, they note that: *'Even where mothers did not have any more support than before, they had changed their attitude. They were proactive in seeking respite and appeared more generally to have a new approach to **coping** with demands.'* (emphasis added, McAuley, Knapp, Beecham, McCurry and Steed, 2004, page 36).

Coping with parenting (technical term: parenting self-efficacy; full definition: the belief that one has the ability successfully to manage the tasks and other aspects associated with being a parent) therefore became the chosen construct to be developed into a measure of impact.

Section 6 Selecting a measurement approach

6.1 Introduction

In parallel to the work undertaken to identify an appropriate construct to capture the overarching impact of Home-Start, the team also undertook background work to inform the selection of a measurement approach. This comprised a focused review of relevant literature, and a selective consultation among other researchers with relevant experience.

As experienced evaluation practitioners, more used to designing large-scale evaluations with complex aims and multi-method approaches designed to be implemented by professional researchers, the challenges presented by this project were new and refreshing. Here, we were not attempting to develop a full-scale evaluation methodology that would capture multiple aspects of a service's functioning in detail. We were, by contrast, aiming to develop a simple, fit-for-purpose tool to support Home-Start to assess the effectiveness of their work at a high level, on an ongoing and routine basis, and at minimal cost. If shown to be successful after testing in the field, in principle the same methods could then easily be adapted to meet the needs of other similar family support services.

The ideal core design elements of the future measure were already established at the outset. We wanted to develop a measure with the following features:

- Valid as a measure of the target construct (coping with parenting, or parenting self-efficacy)
- Quantifiable, and sensitive enough to capture change over time
- Brief and simple to understand
- Low-cost to administer and analyse
- Easily administered by Home-Start staff as part of their routine engagement with parents without disrupting the normal flow of work or relationship-building
- Acceptable to parents and consonant with the service's positive and non-judgmental ethos
- Suitable for self-completion by parents with assistance only in rare cases
- Based on empirical and methodological evidence relevant to measurement science, and as robust as possible, including having been pre-tested, and preferably normed against other relevant populations⁴

Below, we set out the results of our enquiries into the methodological issues that would have to be taken into account in developing such a measure. We summarise what the literature tells us about the advantages and potential limitations of such a scheme, and outline our conclusions about its overall scientific validity and practical feasibility.

⁴ That is, that the measure should have been used in previous studies, and data for comparison should be available from studies of other relevant populations or groups of parents

6.1 Single-perspective self-reporting

Ideally, robust evaluation of any service should use multiple methods and include both investigator-led and respondent-led measures. *Multiple methods* means combining qualitative and various forms of quantitative techniques to gather data. *Investigator-led* means that assessments are made by independent persons, generally professionally trained to do this work, either by observing behaviour or reviewing service users' responses to questions. *Respondent-led* means that the service users themselves, or family members, or service workers, or all of these, provide the assessment and these are accepted as valid in their own terms (i.e., are taken at 'face value'). Sometimes these measurement approaches are combined by constructing a composite score incorporating both forms of measurement, as was used in the assessment of parenting used in Deković et al's study of Home-Start (Deković et al, 2010).

The advantages of having multiple perspectives on how a service is working are obvious: a more 'rounded' view of the situation can be obtained and the biases that might apply to any one respondent's perspective can be balanced. A disadvantage is that different measurement approaches assessing the same construct can produce different results, as can different respondents or observers. Within the literature assessing the impact of parent training on outcomes for children, for example, parents' own reports of changes in their children's behaviour tend to produce evidence of greater service impact when compared to independent observers' reports, possibly due to the fact that parents are biased in favour of expecting change (Maughan, Christiansen, Jenson, Olympia and Clark, 2005). Obtaining multiple perspectives is also relatively time consuming and costly.

In contrast, self-reports by respondents are relatively quick and cheap to obtain. They typically take the form of questionnaires in which participants provide ratings of themselves in relation to a series of questions or statements. However, self-reports are subject to one important form of bias that can undermine the validity of research findings, known as '*social desirability response bias*' where a respondent may intentionally or unintentionally misrepresent their ratings in order to be more consistent with what they believe to be social norms (Foddy, 1993). This can involve 'faking good': downplaying or underreporting undesirable thoughts and behaviours, or the reverse, 'faking bad' (for example to present a more compelling case for receiving services or treatment from a provider). Self-report measures also provide no clue as to the standard by which people may be evaluating themselves. There are likely to be differences in the standards by which one person judges themselves in comparison to another, and these standards may vary over time.

Finally, and very importantly in the context of an evaluation of a service, internal standards may change as a result of participating in an intervention, in which case it is not clear whether a changed self-report, or a self-reported change, is being detected (Brown and Burrows, 1992; McMahon and Metzler, 1998; Howard 1980). Uncontrolled pre- and post-test designs are thought to be especially prone to this problem of 'response-shift' (Howard 1980), which can deliver results so that a problem appears *worse* after intervention. This happens when respondents' initial self-appraisals become less positive the more they reflect upon, learn about, or receive support with, a specific issue.

Despite these drawbacks, self-report measures remain popular in evaluation science because of their convenience and low cost, and because - in the end - there is no substitute for obtaining reports 'straight from the horse's mouth'. However, because of their greater propensity to suffer from various forms of bias, careful attention must be paid to the design of question wording and to the method of administration. There are now well-established safeguards against some of the disadvantages, including question wordings that 'give permission' to a wide range of views and behaviours, and methods of administration that preserve maximum confidentiality and encourage frankness. If response-shift bias is suspected, a 'retrospective pre-test' may be applied whereby respondents are asked to re-appraise their initial responses at the post-test point, in the light of new learning and insights they may have acquired (Campbell and Stanley, 1963; Lamb, 2005). For the purposes of this project, self-report measures were deemed to sit well with the requirement for a simple, brief measure. In comparison to investigator-based measures, they are also more consistent within Home-Start's ethos of non-judgemental, non-intrusive service provision. We therefore concluded that although self-report methods would have some limitations, they were the best fit for our purposes.

6.2 Single-item versus multiple-item measures: reliability, validity and responsiveness

One of the key factors likely to influence the choice of self-report measure for use in an evaluation is the length of a measure and the time it takes to complete. Most measures contain multiple items, and the number can vary considerably. Less typical are single-item measures that attempt to capture a participants' position in relation to a construct using a single, global question or statement. A (hypothetical) example of a single item question at its shortest is: *Are you coping with parenting, yes or no?* A more developed format would be: *On a scale of one to ten, where one is 'not at all' and ten is 'all the time', how well are you coping with parenting?* An example of a multiple item measure, by contrast, would be the well-known *Parenting Stress Index* (PSI, Abidin 1983, 1995) which poses 120 (original version) or 36 (short form) questions about different aspects of stress and parenting and a scale score is derived by summing the responses to the constituent items.

Regardless of a measure's length, in order to be robust, a measure must demonstrate acceptable levels of so-called 'psychometric properties' such as reliability and validity. *Reliability (or stability)* means the measure captures the same thing in the same way consistently over time, and for scales, that the items that comprise it are appropriately related to one another. *Validity* is defined as the ability of a measure to assess what it says it measures. Measures that are used to assess change over time, such as improvement between time points prior to and after using an intervention also need to demonstrate sensitivity to or the ability to detect change, known as '*responsiveness*' (Terwee, Dekker, Wieringa, Pummel and Bossuyt, 2003).

6.2.1 Reliability and validity

Single item or very short measures are often suspected of being less reliable, less valid and less responsive than more complex and length multiple-item measures. However, there is evidence that single-item measures can have acceptable reliability and validity, although

this evidence is derived predominantly from health and organisational psychology rather than the field of family support research and evaluation. Bowling (2005) reviewed the use of single-item assessments in the measurement of health status, quality of life and health-related quality of life. Single, global questions asking respondents to rate their health as 'excellent, good, fair or poor' have been widely used in this field as they can capture diverse aspects of health and illness in a single summary item. Bowling reports that these measures have a long tradition of use, and generally have been found to be reliable and valid. However, she points out that some variation in reliability has been found depending on the position of the item relative to other questions in the survey, and also in relation to the standard that the respondent is using as a benchmark of acceptable health.

Further evidence comes from the fields of organisational psychology and child psychiatry. Wanous, Reichers and Hudy (1997) carried out a meta-analysis, statistically synthesising the results of 28 studies in which the relationship between single-item measures and multiple-items assessments of job satisfaction were examined. They found that scores on single-item measures were strongly associated with scores based on multiple-item scales; in other words, that single-item measures can 'stand in for', and do as good a job of measuring key constructs as multiple-item scales. A similar finding in a field more relevant for our purposes is reported in relation to the well-known and widely used *Strengths and Difficulties Questionnaire* (SDQ; Goodman, 1997), a 25-item questionnaire assessing children's emotional and behavioural problems. In a paper published in 1999, Goodman reports that a single item asking parents about the overall impact of their child's difficulties on the family performs just as well in distinguishing between children with and without clinically significant emotional and behavioural disorders as the full 25-item scale, which taps a range of specific outcomes (Goodman, 1999). This of course is a very interesting finding providing encouraging support for the principle of using overall, global measures of impact.

6.2.2 Responsiveness

In respect of responsiveness (ability to detect change over time), however, one limiting factor is that short and single-item measures may be less sensitive. This means that larger sample sizes may be required to detect group differences in comparison to the sample sizes required when using more elaborate measures (Diehr, Chen, Patrick, Feng and Yasui, 2005; McHorney, Ware, Rogers, Raczek and Lu, 1992).

In Bowling's (2005) review of single, global health assessment questions, she concludes that a multi-item scale, even a relatively brief one of between five and ten items, is more sensitive to changes in respondents' health status than a single item. This is supported by a study comparing a five-item and single-item assessment of health status carried out by Diehr and colleagues. Although the responsiveness of the single-item measure was acceptable, it was lower than that of the five-item measure, and the authors concluded that a larger sample size would be required to detect statistically significant change when a single-item measure is used (Diehr et al, 2005). The format used for response options is also important. Simple binary response scales that use a 'yes-no' or 'true-false' response format may be too crude to detect subtle changes among respondents, and scale formats of up to seven points (known as Likert scales) are generally thought to be more sensitive to detecting change (Hyland, 2003).

However, although a small number of items can result in greater responsiveness of a scale when compared to a single item, it has also been suggested that responsiveness is reduced if there are *too many* items, because the greater the number of items, the greater the possibility of inclusion of items insensitive to change (Brown and Burrows, 1992; Fitzpatrick, Fletcher, Gore, Jones, Spiegelhalter and Cox, 1992). This may arise because some items are simply not applicable to some respondents (McMahon and Metzler, 1998). It can also arise when a questionnaire assesses a relatively stable factor or one not amenable to change by the intervention. Hence the importance of aligning intervention goals and the constructs being measured (Hyland, 2003; McMahon and Metzler, 1998; Vermeersch, Lambert and Burlingame, 2000). What constitutes the 'correct' number of items to optimise responsiveness is therefore uncertain. In longitudinal studies assessing quality of life, Hyland (2003) recommends 'short' scales of between one and forty items. In other fields of study, including family support, the issue of optimal scale length for assessing change is not generally specified.

Finally, the characteristics of the population studied and the rate at which a given condition is found in the group being measured will also influence a scale's responsiveness. In particular, an effect known as '*ceiling*' or '*floor*' effects may reduce the responsiveness of a measure. This occurs when most people score close to the top, or to the bottom, of a scale, leaving no space for detection of any change at re-test. These effects can arise if the measure was designed for use with samples whose characteristics differ from those on whom it is being tested. However, as well as reflecting genuine lack of diversity within a sample, a ceiling or floor effect may also arise when respondents are unwilling to admit to a particular view or behaviour, in which case these effects become correlated with social desirability response bias.

Responsiveness is therefore not simply a characteristic of a measure (in terms of number or items, response format, and relevance of content). It also varies as a function of the population being assessed (Hyland, 2003), and is likely to be compromised when measures are 'imported' from studies on one type of population to another without proper testing.

6.3 The final selected approach

Weighing the various factors noted above, we concluded that there is not, in fact, compelling evidence against the use of brief, single-item measures provided they are well-constructed and appropriate to the circumstances. There may also be some important advantages. These include, not least, that if one short measure can capture relevant data as well as a longer and more elaborate one, it is wasteful of effort and burdensome on respondents to do otherwise. Wanous and colleagues note the importance of purpose and context in considering the use of a single-item measure. They advise that such measures can be appropriately used 'when the research question implies their use or when situational constraints limit or prevent the use of scales' (Wanous et al, 1997, p50). For our purposes, there are certainly situational constraints on Home-Start scheme Co-ordinators and parents and the single-item approach or very brief measure would seem potentially appropriate.

However, the research question in this instance concerns the assessment of change among service recipients over time rather than assessment at a single point in time. A critical issue for our purposes here, therefore, is whether a single-item measure or very brief measure would be sensitive enough to capture change among Home-Start parents in comparison to a more elaborate measure. In other words, the 'responsiveness' of the measure is the determining factor, which could only be assessed by testing the measure in the field. We describe this process further in Section 9.

Section 7 Selecting the measures to be tested

7.1 Introduction

In order to meet the requirement for a measure assessing an overarching construct that could be assessed in a brief, simple way, we reviewed the literature concerning theories and constructs relevant to family support, as well as literature concerning measurement issues. In this section of the report we draw together the conclusions from this process in order to identify individual measures for testing in the field.

7.2 Criteria for selecting among measures

The recommended characteristics of a measure for our purposes here were that it should:

- Be an acceptable construct to capture the global, overarching impact of Home-Start's work as recognised by staff and volunteers: ie, parental coping or parental self-efficacy, or a construct that is closely related to it
- Use self-report methods rather than observer-rated or interviewer- assessment methods, due to constraints on time and budget.
- Contain as few items as necessary to retain good psychometric properties, not only for the sake of brevity but also in order to optimise responsiveness of the scale
- Use a response scale that is multi-point rather than binary to enhance responsiveness of the scale and its ability to detect change in realistic sample sizes (the definition of realistic may vary from one study to another)
- Have been tested with similar families to those using family support services such as Home-Start, to demonstrate that it is acceptable and appropriate for this type of sample, and to avoid ceiling or floor effects.

The choice of outcome measures for more extensive evaluations would be guided by different criteria that would not necessarily overlap with the above. Given weaknesses in the evidence base in this field, some of our conclusions are more tentative than others. For example, the need for reliability and validity of a measure are well established psychometric requirements, but the number of items needed within a measure in order to optimise its responsiveness is less certain. The selection of a measure that meets all of the above recommendations might not be possible, and could involve prioritising some criteria above others. The various limitations noted in Section 6 must be acknowledged, and the functionality of any chosen measure would remain unknown until more substantial field testing, measuring change over time, has been completed.

7.3 Measures of parenting self-efficacy

Self-efficacy in the context of parenting is not a consistently utilised term. In the present discussion, we use the term 'parental efficacy' loosely in order to take into account the variations in terminology used within the literature. Subtle differences in definitions of the construct are offered by different authors. Constructs such as parental self-esteem, self confidence or competence, for example, tend to be used synonymously with self-efficacy, even though they are argued to be conceptually distinct (De Montigny and Lacharité, 2005).

Bandura regarded self-efficacy as domain or task specific, and did not use the construct in a 'global' sense. Some authors offer definitions of constructs that are consistent with Bandura's (1997) definition of parental self-efficacy, but label them differently, such as 'parental self-agency' (Dumka, Stoerzinger, Jackson and Roosa, 1996). Others have adapted the construct to be more overarching and less domain-specific. A significant advantage of the more global measures of parental self-efficacy is their broad applicability to parents whose children may span a relatively wide age range (Coleman and Karraker, 1998). This is important in the context of family support services such as Home-Start where the focus is not exclusively on parents of children within a particular age group such as infants, but on parenting more broadly. Given that our primary purpose here is to identify a measure that demonstrates Home-Start's overarching impact, a global measure is appropriate.

Examples of commonly used global parental efficacy measures include the *Parental Sense of Competence Scale* (Gibaund-Wallson and Wandersman, 1978, cited in Johnson and Mash, 1989); the Competence subscale of the *Parenting Stress Index* (Abidin, 1983); and the *Parenting Self-Agency Measure* (Dumka et al, 1996), which are all US-developed measures. UK-developed scales include the *Tool to Measure Parental Self Efficacy* (Kendall and Bloomfield, 2005) and a single-item measure of *Coping with Being a Parent* developed by Gbate and Hazel (2002).

We therefore identified five measures that met most of the criteria listed in section 7.1. These were further reduced to three through a process of weighing up the advantages and disadvantages of each. As part of the consultation exercise described in Part One, Home-Start Co-ordinators were invited to review the measures and give feedback (focusing on how relevant, how easy to understand, and how practical and appropriate the questionnaires were in the context of an assessment with a family). They concluded that two of the three measures were appropriate, and one item extracted from the third measure was also considered as a potentially useful item for pilot testing. The rejected measures and reasons for rejection are briefly outlined in Appendix 2. We now describe the measures that were selected as the best fit for our purposes.

7.3.1 The *Parenting Self-Agency Measure* (PSAM)

Box 2 The *Parenting Self-agency Measure*

Please put a tick in the box that best describes you:	Almost never or never	Once in a while	Sometimes	A lot of the time	Almost always or always
I feel sure of myself as a mother/father					
I know I am doing a good job as a mother/father					
I know things about being a mother/father that would be helpful to other parents					
I can solve most problems between my child and me					
When things are going badly between my child and me, I keep trying until things begin to change					

The Parenting Self-Agency Measure (PSAM; Dumka et al, 1996; box 2) is one of the briefest measures of self-efficacy, involving only five items that are all phrased positively, and uses a five-point response format ranging from 'Almost never or never' to 'Almost always or always'. It has been widely used in research both as a cross-sectional (one-off) measure to assess baseline levels of parenting self-efficacy and their relationship with parenting practices (for example, Coleman and Karraker, 1998) and also as a measure of change over time to assess the effectiveness of parenting support interventions. For example, Letarte, Normandeau and Allard (2010) used the PSAM to measure change as a result of participating in the Incredible Years programme amongst a sample of parents in the Canadian child protection system, although results in that study showed no significant change over time. In our consultation with Home-Start Co-ordinators, the brevity of the scale was appealing as it was seen as a minimally intrusive measure that could fit relatively easily within assessment visits. The use of positive wording was seen by Co-ordinators as an important aspect of the scale because of its fit with the positive perceptions of parenting that Home-Start wishes to promote. It is also consistent with Bandura's belief that efficacy should be measured using a positive, 'can do' phrasing of items (Bandura, 2006).

The scale has been psychometrically tested among Anglo Americans, and Spanish-speaking immigrants, within the US where it was developed, and has been found to have good internal reliability and construct validity (Dumka et al, 1996). The scale has also been used with parents in the UK, among families attending health visitor clinics (Whittaker and Cowley, 2006). The UK study assessed the internal reliability of the measure, the construct validity, and the test-retest reliability of the measure over a two-week period and found the levels of reliability and validity to be acceptable. Feedback comments from parents in the study indicated that most found the measure to be acceptable.

One drawback of this scale is that two of its five items ask about parenting of a 'target' child. Since Home-Start does not work with specific children in the family, we adapted the scale in order to capture self-reported parental efficacy in relation to parenting of all children within the family, which we did in consultation with the author of the scale (Dumka, personal communication). Therefore, parents were instructed to consider all of their children rather than an individual child in responding to these items.

7.3.2 'Coping with Being a Parent': from the national survey of parents in poor environments – modified into the *Parent Coping Scale*

Box 3 The Parent Coping Scale

Most people find being a parent has its ups and downs. Taking everything into account, which of these statements best describes how you are coping with being a parent these days?

Circle one number:

- 1. I feel I am not coping at all these days*
- 2. Most of the time I feel I am not coping very well*
- 3. Sometimes I feel I am coping but sometimes things get on top of me*
- 4. Most of the time I feel I am coping pretty well*
- 5. I always feel I am coping really well – things never or hardly ever get on top of me*

The other scale selected for inclusion measured 'coping' with parenting in its most global 'domain general' sense. 'Coping with Being a Parent' was originally developed by Ghate and reported in Ghate and Hazel (2002). This measure was designed for use in a large scale nationally representative survey of stress and coping among parents in Britain living in poor neighbourhoods funded by the British Department of Health. It is a single-item measure that has now been used in many large studies of families in the UK in addition to the original study of parenting in poor environments. A modified version concerning 'coping with your child in the past month' was also used in the national evaluation of the Youth Justice Board's Parenting Programme (Ghate and Ramella 2002). The question was originally designed as a cross-sectional indicator to establish levels of parenting need and the relationship of levels of coping to other indicators of risk for parenting including parent mental health, child behaviour and so on. However, it was later utilised as a measure of change over time in a large scale quasi-experimental longitudinal cohort study of parents living in disadvantaged areas as part of the national evaluation of *On Track*, a government family support and children's services initiative. In that study, statistically significant increases in levels of coping over a one-year follow-up period were reported both for parents using the On Track services, as well as for those resident in the On Track service area, compared to parents in a matched comparison group (Aye Maung, Parfremment and Tipping (2008); Ghate, Asmussen, Tian and Hauari 2008 p183). In the national evaluation of the Youth Justice Board's Parenting Programme, a pre and post-test (uncontrolled) design also showed significant improvements in a similar measure of coping after participating in a nine-week group-based parenting support intervention.

There is considerable appeal in the use of single-item measure with the families with which Home-Start works, as it places least possible burden on practitioners and families, and is minimally intrusive to the service's work. Co-ordinators praised this scale for being 'quick', 'simple' and 'easy to use'. This particular measure also has the advantage of assessing parents' self-reports without reference to a 'target' child. The measure as originally designed involves four statements: 'I am coping pretty well with being a parent; things rarely get on top of me'; 'Sometimes I feel I'm coping well, but sometimes things get on top of me'; 'I hardly ever feel I'm coping well'; and 'I'm not coping at all these days'. Parents are asked to choose one statement that best reflects how they currently feel they are coping with parenting. Although this item measures coping, it does not do this in relation to any particular coping strategy, unlike most measures of coping. Instead this measure uses the term 'coping' globally. Self-efficacy is argued to be a component of cognitive appraisal that influences coping (Bandura, 1997), as discussed previously. Hence parents with high self-efficacy are likely to see themselves as coping well.

In previous studies, differences in coping among parents were found to be strongly associated with a number of parent and child difficulties including a tendency to parental depression and having a child with behavioural or emotional difficulties, as well as with the need for and use of different forms of parenting support (Ghate and Hazel, 2002; Ghate, Asmussen, Tian and Hauari, 2008). Although this indicates the measure has good validity, the measure as originally designed had never been assessed for test-retest reliability. The distribution of responses to this item in the original survey also showed some indications of a ceiling effect, possibly reflecting not a genuine ceiling effect but a social desirability bias, in that that only 2% of parents rated themselves on the last two statements concerning coping badly or not at all. Hence for our purposes here, a fifth statement was added in

order to provide a greater potential spread of scores, to avoid ceiling effects and enhance potential responsiveness of the measure. This amendment also involved a slight adjustment to the wording of the remaining items. The resulting modified scale was then renamed as the *Parent Coping Scale (PCS)*; box 3.

7.3.3. The *Enjoyment of Parenting* measure

Box 4 The *Enjoyment of Parenting* measure

Please put a tick in the box that best describes you					
I enjoy being a parent	Almost never or never	Once in a while	Sometimes	A lot of the time	Almost always or always

One further item was included among the measures to be tested. This was a single question concerning parents' enjoyment of parenting, which regularly appears in a wide range of research studies on parenting and on family support, including for example the Competence subscale of the *Parenting Stress Index* (Abidin 1983, 1995); see box 4.

Satisfaction with parenting has been found to be related to parental self-efficacy (Gilmore and Cuskelly, 2008), as it is argued that people are more likely to be efficacious in areas that they find rewarding (Bandura, 1982). Thus, items asking about satisfaction with parenting have been included in several scales measuring the specific construct of parenting efficacy. The *Parental Sense of Competence* scale, for example, contains a subscale assessing satisfaction in addition to self-efficacy, and the combined score on both satisfaction and self-efficacy are sometimes used in studies to report on self-efficacy. In the previously mentioned study of parenting in poor environments (Ghate and Hazel 2002), parents were asked to respond, within a self-completed booklet, to the statement 'I really enjoy being (child's) parent' for each child in their family family, on a four-point scale (all of the time to never).

When Home-Start Co-ordinators reviewed this item in the consultation meeting, they were extremely positive about this question, with some believing that a core aim of the service was, in fact, to increase enjoyment of parenting, or, as Home-Start's promotional literature frames it: to put "the fun back into family life" (Home-Start 2010a). Given this, and its reported close association with parental self-efficacy and parental satisfaction, it was decided to include a single item assessing this construct. Hence, the statement '*I enjoy being a parent*' was included in our selection of measures, and rated on the same five-point response scale as the items in the *PSAM*. The reliability and validity of this single item is not known, although very similar items from other measures are included in many scales that have been shown to be robust.

These three selected measures were combined into a 'Parent Questionnaire' (shown in Appendix 1) and this was used for pilot testing with Home-Start families.

Section 8: Pilot testing of the selected measures

8.1 Introduction

Having selected three alternative measures – the *Parenting Self-Agency Measure*, the *Parent Coping Scale*, and the *Enjoyment of Parenting* measure – and combined these into a short questionnaire (see Appendix 1), we tested the measures in a pilot study organised, supervised and implemented by Home-Start Northern Ireland between October 2010 and September 2011. The aims of the pilot study were not to generate representative data on the actual impact of Home-Start services in the participating local Home-Starts. Rather, the aim was to assess both process and measurement issues relevant to the attempt to develop a measure of overarching impact for use in future attempts to assess the effectiveness of Home-Start's work. The pilot was therefore designed to explore the feasibility and plausibility of each of the measures, and its specific objectives were:

1. To explore the practical issues in administering questionnaires and collecting data from the perspective of Home-Start Co-ordinators
2. To test the acceptability of the process in general to parents
3. To test the relevance, understanding and acceptability of each of the three alternative measures to parents
4. To explore the responses provided in relation to the robustness of each measure, including its psychometric properties, and to make inferences based on the patterns of response about the likely suitability of each measure for the purpose of assessment of overarching impact of the service

The piloting of the measures could not have been completed without the considerable efforts of the regional team in Northern Ireland, and the local scheme Co-ordinators. It was perhaps a particularly robust test of the methods that the piloting period also coincided with a period of intense difficulty for Home-Start (as for the family support field and voluntary sector providers generally). During this time, a substantial programme of austerity measures was being introduced by central and local government in Northern Ireland as elsewhere in the UK, in response to the global economic crisis. For Home-Start in Northern Ireland, this led to reduced funding for five of their 24 schemes and the merger of two others during the study period.

8.2 Pilot data collection procedures

8.2.1 Repeated measures design

In order to measure changes in impact over time, a 'repeated measures' approach is required. This requires that service users complete a measure on repeated occasions, including at initiation of service, and again at one or more follow-up points. Follow-up points must be chosen to reflect a sufficient time period over which positive changes might realistically be expected to be observed.

In practice, when a family is first referred into a local Home-Start, an 'initial (assessment) visit' is undertaken by the paid Co-ordinator. This is followed by a 'match visit' when a

volunteer is introduced to the family by the Co-ordinator. Subsequent visits are then made by a volunteer who has been matched with the family, with support and supervision by the Co-ordinator. Generally, volunteer visits are made weekly for the duration of service. Co-ordinators also visit the family to undertake 'review visits' every 10 to 12 weeks. Since the aim of the project was to develop a measure that could be incorporated into Home-Start's business as usual as far as possible, we wanted to gauge how practical it was to administer the measures to parents during their routine initial assessment and subsequent follow-up visits. Co-ordinators were asked to take on this role as part of their initial assessment and routine subsequent progress checks.

We selected three time points for data collection: the initial assessment visit at which Co-ordinators met families and carried out Home-Start's usual assessment prior to volunteers being matched to a family; and two follow-up review visits at approximately 10 week intervals, carried out after volunteers had begun working with families. These are referred to as *Time One* or *Baseline* (first contact with the service), *Time Two* (approx 10 weeks after baseline) and *Time Three* (approx 20 weeks after baseline) in the rest of this Section. The whole study period from initial to second follow up visit was designed to last 20 weeks, but in practice, because of natural variations in when visits actually took place, the total 'follow-up period' varied across respondents from 15 to 40 weeks, and was on average 23 weeks.

8.2.2 Participating schemes and number of participating respondents

The pilot data were gathered by Home-Start Co-ordinators based in Northern Ireland, and all schemes based there were invited to take part. Among 24 schemes in the country, 12 agreed to collect data from all families who initially commenced access to the service between October 2010 and April 2011. In total, 88 families joined Home-Start services in the 12 local Home-Starts during this period, fewer than usual due the constraints described above in 8.1. Of these 88 families, 76 parents agreed to complete an initial (Baseline, Time One) questionnaire. This response rate of 86% is extremely respectable by routine survey and evaluation standards.

8.2.3 Data collection by Co-ordinators

Co-coordinators were briefed at a meeting prior to data collection on how to recruit, introduce and implement the study, and supported throughout by a project manager based locally. She collated completed paper questionnaires and returned these to head office where data were entered into Excel, and subsequently exported to SPSS (a data analysis package) for analysis by the research team.

Co-ordinators were asked to invite all families accessing the service for the first time between October 2010 and March 2011 to take part in the pilot project. In practice, and reflecting Home-Start's usual way of working, the main carer (generally a mother) was the person invited to participate. During initial assessment visits to these families, Co-ordinators therefore explained the purpose of the evaluation project to parents and also gave assurances about confidentiality, anonymity and the right to withdraw from the pilot study at any time without adversely affecting the service they were receiving from Home-Start. Parents were given a short information sheet that summarised this (see Appendix 3), and were also asked to sign a consent form if they agreed to take part.

At the initial assessment visit, Co-ordinators routinely gather family demographic details and other information that allows them to understand the presenting needs of newly referred families. During the course of the pilot study this process was being re-configured as part of a wider data collection exercise known as 'MESH' (Monitoring and Evaluation of Service for Home-Start). This caused some issues for the pilot study, as some of the new questions in MESH also use the term 'coping'. This led to some confusion for Co-ordinators about the differing purposes of *assessment of need* and *assessment of impact* and caused some to comment, perhaps not unfairly, that some of the impact questions in our short questionnaire seemed duplicative.

At the two subsequent review visits, each approximately ten weeks apart, the Co-ordinators asked parents to complete the three alternative impact measures again. At the end of the study period, Co-ordinators also provided data regarding the families' use of Home-Start such as the duration of involvement with the service, the number of volunteer visits, and reasons for contact with the families coming to an end, if this had occurred.

Finally, feedback from Co-ordinators was a vital part of the pilot study. We wanted to understand the practicalities of gathering data from parents during Home-Start visits, and to obtain insights into parents' acceptance of the questionnaires and their approach to completion of the questionnaire from the perspective of Co-ordinators. An open-ended process feedback questionnaire was designed specifically for this purpose, and is shown in Appendix 4. Findings and implications of the feedback are discussed below in Section 9.

Section 9: Findings of the pilot study

9.1 Participating families

Of 88 parents approached, 12 (14%) declined to take part. This refusal rate is similar to the rate of 10% reported in a recent study of Home-Start by Deković and colleagues (2008a) and is a respectable response rate by evaluation research standards. Reasons for refusals are shown in table 1.

Table 1. Reasons given by parents for declining to take part in the pilot study

Reason for refusal (n=12)	No.
Situational constraints e.g. nursing baby, toddler crying	3
Parent incapacitated by physical pain or depression	2
Parent only requires very short-term support	1
Parent is ante-natal	1
Parent doesn't speak English	1
Parent reluctant to provide information/"Not my sort of thing"	2
No reason given	2
Total	12

All 76 participating parents were mothers, ranging in age from 17 to 45 years with an average age of 31 years. Over half described themselves as British (54%), and over a third as Irish (37%). The remainder were either 'other white' (5%) or African (3%). Most were married or co-habiting (71%) and around two-thirds (67%) were not in paid employment, although a similar proportion lived with another adult in the household who was working (61%). Almost half (46%) lived in privately owned accommodation, and a third (34%) lived in privately rented accommodation, while the remaining fifth (20%) lived in local authority or housing association accommodation. The number of children in the household varied from 1 to 7, with an average of 3. Sixteen percent of families had at least one child with a disability or special need, and 11% had at least one child who was on the child protection register or the subject of a childcare plan.

Parents who declined to take part did not differ significantly from those who agreed to take part with respect to age, ethnicity, marital or employment status, family size or number of children. This, combined with the high response rate obtained, shows that at least in principle, the pilot study was based on a group of service users who were representative of the population of new Home-Start service users during the research period.

9.2 Sources of referrals

The largest source of referrals to Home-Starts in the pilot study were health visitors, who referred just over half of parents taking part (53%). Other main sources of referral included social services (referring 12% of participants), and other health and social service personnel (9%). Parents referred themselves to the service in 16% of cases. A breakdown of the

sources of referrals is provided in table 2. (Figures may not add up to 100 for percentages due to rounding.)

Table 2. Referral source

Referrer (N=76)	%	(n)
Health visitor	53	(40)
Self referred	16	(12)
Local health and social services (referrer not specified)	12	(9)
Sure Start	4	(3)
Nursing specialist (e.g. midwife, psychiatric nurse)	4	(3)
Voluntary sector agency	4	(3)
G.P.	1	(1)
Not recorded	7	(5)
Total	100	(76)

The referral pattern is very similar to the figures reported for Home-Start Northern Ireland in Home Start's annual statistics for 2009/10 (Home-Start, 2010b), and again, gives us confidence regarding the representativeness of the pilot sample.

9.3 Baseline (Time One) results at initial assessment visits: 76⁵ parents

9.3.1 Baseline results on the *Parenting Self Agency Measure (PSAM)*

The *PSAM* is a multiple item scale. It consists of a series of 5 statements about parents' beliefs about aspects of parenting, tapping the construct of parenting 'self-agency'. It is prefaced by the following explanation, which focuses on the parent's contemporaneous perceptions: *The statements below describe feelings and thoughts about being a parent. Please tell us how often you feel or think like the statement, thinking about your current situation.* Self-agency is a construct more or less indistinguishable from 'self-efficacy', discussed in detail previously in Sections 5 and 7. Each statement can be answered on a scale of frequency from '*almost never or never*' (score = 1) to '*almost always or always*' (score = 5). The scale was developed by Dumka et al (1996) for use in the United States but has previously been used by Whittaker and Cowley (2006) in the UK in a study of parents accessing routine health visitor services in the community study⁶.

Table 3 shows the distribution of *PSAM* scores (n=75, 1 missing). At the Baseline point, scores ranged from 5 to 25, the same as the minimum and maximum scores possible on this measure. The *PSAM* mean average score for Home-Start parents was 18 (s.d.=4.3). Table 3 shows that on nearly all of the items, most parents scored themselves on the more positive end of the scale, i.e. from the mid-point of the response scale towards the upper end of the scale. One exception to this was item number 3 concerning '*I know things about being a*

⁵ Data were missing for one parent

⁶ Health visitor services in the UK are a universal service provided by the National Health Service to all parents of new born infants and covers the entire population.

mother/father that would be helpful to other parents'. For the latter item, scores were more evenly distributed.

Although the distribution of responses indicates that parents were rating themselves towards the more positive end of the scale, the average *PSAM* score reported among Home-Start parents is only slightly lower than the average score on the same scale as reported by Whittaker and Cowley (2006), which had a similar-sized sample but included a more socially mixed group that had not necessarily sought referral to support services. This study found an average score of 19.9.

Table 3. Distribution of responses at Baseline for the *Parenting Self Agency Measure (PSAM)*

PSAM items 1 to 5 N=75	Almost never or never % (n)	Once in a while % (n)	Sometimes % (n)	A lot of the time % (n)	Almost always or always % (n)
1. I feel sure of myself as a mother/father.	5 (4)	8 (6)	29 (22)	25 (19)	32 (24)
2. I know I am doing a good job as a mother/father.	5 (4)	5 (4)	35 (26)	33 (25)	21 (16)
3. I know things about being a mother/father that would be helpful to other parents.	13 (10)	17 (13)	37 (28)	20 (15)	12 (9)
4. I can solve most problems between my child ⁷ and me.	4 (3)	7 (5)	29 (22)	37 (28)	23 (17)
5. When things are going badly between my child and me, I keep trying until things begin to change.	4 (3)	1 (1)	21 (16)	36 (27)	37 (28)

9.3.2 Psychometric properties of the *Parenting Self Agency Measure* at Baseline

General psychometric properties for this scale were reported in Section 7. We also assessed the internal reliability of the *PSAM* in this study. This test, which generates an 'alpha coefficient', assesses how much the individual items within the measure are assessing a single concept and are appropriately related to one another. A higher alpha coefficient figure indicates a greater degree of internal consistency. The result for the *PSAM* was 0.85 (where 1 is the maximum possible), indicating good internal consistency.

However, there is some evidence from our data that either a ceiling effect, or a social desirability response bias ('faking good') may have been operating in this study. A large proportion of parents reported themselves to have high self-agency, with 57% of parents

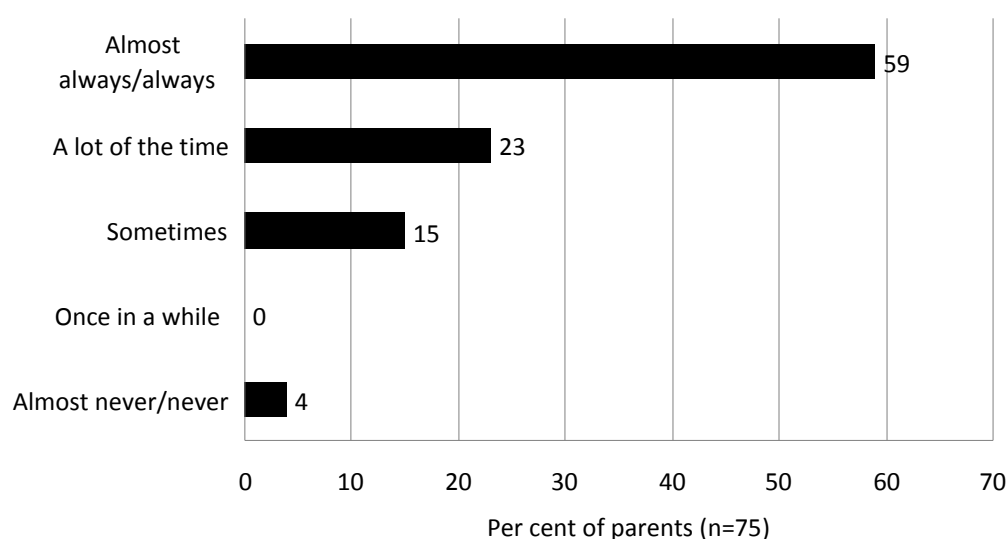
⁷ Note that we issued an instruction to answer this question in relation to all children in the family, if more than one; see Section 9.5.5.

saying they always or usually ‘felt sure’ of themselves, 54% feeling they were ‘doing a good job’, 60% ‘able to solve problems’ and 73% able to achieve positive change. Given that this was a group of parents seeking support with parenting, these figures may perhaps seem implausibly positive. The fact that they are very similar to figures for a general population sample receiving routine health visitor services also seems surprising, and suggests social desirability bias. Below in Section 9.7 we report on feedback from Co-ordinators who administered the questions, which tends to confirm this speculation.

9.3.3 Baseline results on the *Enjoyment of Parenting* measure

This measure is a single-item scale. It appears in a number of longer scales used to measure aspects of parenting in cross-sectional studies (e.g Abidin, 1983, 1995; Ghate and Hazel 2002). We used it in a single statement form, answered on the same frequency scale as the *PSAM*. Scores for Home-Start parents ranged from 1 to 5, which are the minimum and maximum scores possible. The mean score was 4.3 (s.d.=1, n=75, 1 missing). As many as 59% of parents said that they *always or almost always* enjoyed being a parent. Four percent of parents reported that they *never or almost never* enjoyed being a parent. This means that the overwhelming majority of parents (82%) stated that they enjoyed being a parent most or all of the time. Figure 1 shows the distribution of responses.

Figure 1. Responses to the ‘Enjoyment of Parenting’ measure at Baseline



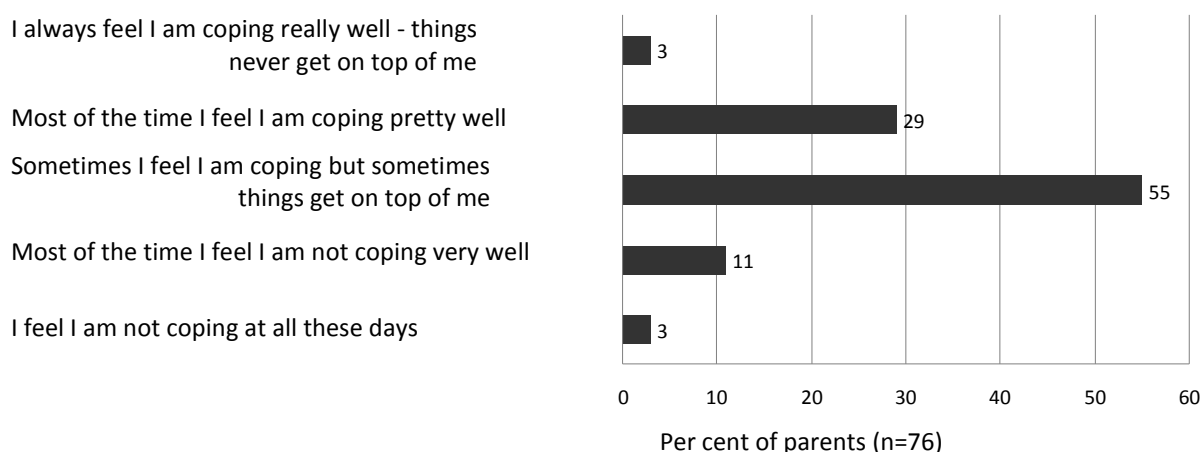
9.3.4 Psychometric properties of the *Enjoyment of Parenting* measure at Baseline

Reliability (test-retest) data are not available for this measure, and internal consistency tests are not applicable as it is single item scale. However, as with the *PSAM*, our data suggest some evidence of a ceiling effect, possibly related to a social desirability response bias.

9.3.5 Baseline results on the *Parent Coping Scale*

As noted in Section 7, this measure is a single item, global (domain-general) coping scale, adapted from a measure of Coping with Being a Parent (Ghate and Hazel 2002). We modified the original scale statements for this study by including an extra scale point in an attempt to correct the social desirability response bias strongly evident in the original study where very few parents admitted to substantial problems with coping. The demographic and social characteristics of parents who took part in this earlier national survey were broadly comparable to those of parents using Home-Start, except that that they were not necessarily actively seeking support with parenting.

In the pilot study, parents were asked to select one option, from five statements, that best described how they were coping with being a parent. The question was, as in the original study, prefaced by a 'permission-giving' statement that is overarching in scope and framed in accessible, colloquial language for a British-English speaking population (*'Most people find that being a parent has its ups and downs. Taking everything into account, how well are you coping with being a parent these days?'*) The proportions of parents endorsing each statement are shown in figure 2. Over half (55%) rated themselves at the mid-point of the scale, indicating sometimes they coped and sometimes they did not. Almost a third (29%) felt they were 'coping pretty well' most of the time. When compared to results from Ghate and Hazel's (2002) study, we see that there were similar proportions of parents positioning themselves at the 'sometimes coping and sometimes not' point on the scale, with 53% reported in Ghate and Hazel and 55% among Home-Start families. However, while Ghate and Hazel found only 3% of parents reporting that they hardly ever coped well or coped not at all, the comparison figure for Home-Start families was 14%. In the current study the mean score was 3.2 (s.d.=.8, n=76).

Figure 2. Responses to *Parent Coping Scale* statements at Baseline

9.3.6 Psychometric properties of the *Parent Coping Scale* at Baseline

Internal consistency tests were not required for this measure as it is a single item scale. Encouragingly, at Baseline the distribution of responses in this study suggested that the tendency to social desirability response bias that was strongly evident in the original study using the unmodified measure developed by Ghate and Hazel had been corrected: in the Home-Start study, 14% of parents admitting to regular difficulties appears to be a much more plausible proportion, especially in light of the fact that as active seekers of support, parents in the Home-Start pilot sample would be expected to report some difficulties with coping.

At the time of the Baseline measures, information on the stability of the scale over time points (using test- retest procedures) was not yet available. However, during the course of the project, a separate assessment of the test-retest reliability of the Parent Coping Scale was carried out in order to confirm stability of the measure. Parents attending group meetings in local Home-Starts based in one area of England were asked to complete the PCS on two occasions, one week apart. Results based on a sample of 34 parents show an intra-class correlation coefficient of 0.93, with 95% confidence limits of .86 to .97. This correlation coefficient indicates that the measure shows excellent reliability.

9.3.7 Relationships between the measures at Baseline

Lastly, to ensure that the different measures used in the testing were different to another, yet broadly tapping the same constructs, we assessed the degree of association among the measures using a statistic known as 'Spearman's rho', in which a higher value indicates a stronger association. The results are shown in table 4 below, and all of the correlations were statistically significant at the $p < .01$ level (2-tailed test). This confirms that the measures are assessing areas that are positively associated with one another (or in other words, are conceptually related). This is also consistent with other studies that show a

relationship between parental self-efficacy and role satisfaction, although the direction of any causal relationship is unclear (Jones and Prinz, 2005).

Table 4. Association among measures of parental self-agency, enjoyment of parenting and the Parent Coping Scale at Baseline (n=75)

Measure	PSAM	Enjoyment of parenting	Parent Coping Scale
PSAM	-	.63**	.54**
Enjoyment of parenting	-	-	.38**

**P<.01

9.4 Results from follow-up visits

9.4.1 Follow-up visits

After an initial assessment visit, Co-ordinators matched a volunteer to each family, and volunteers began visits to families at regular intervals. Over three-quarters of families (76%) received their first visit from a volunteer within a month of being assessed by the Co-ordinator, with an average wait of three weeks (s.d.=3.4, n=64, 12 missing).

Co-ordinators aimed to carry out two follow-up review visits to families at intervals of approximately 10 weeks after the initial assessment visit had been carried out. This interval was designed to fit with Co-ordinators' typical pattern of review visits.

During these two review visits Co-ordinators asked parents to complete the measures again. The interval between the initial (Baseline) visit and the first review visit (Time Two) was on average 13.6 weeks (s.d.=5.5, n=49, 2 missing), and ranged from 8 to 32 weeks. The length of time between the first review (Time Two) and second review visit (Time Three) ranged from three to eight weeks, with an average of 10.7 weeks (s.d.=3.5, n=33, 1 missing). The total study period, from Baseline to Time Three was on average 23 weeks long (s.d.=5.7, n=33, 1 missing) and ranged from 15 to 40 weeks.

9.4.2 Families with follow-up data

Of the **76** families providing data at the Baseline, 25 did not go on to provide follow-up data at Time Two. Between Time Two and Time Three, a further 17 families did not provide data. Thus in summary, we received data from both the initial visit and the Time Two visit for **51** families (i.e. 67% of the original sample). We received data from both the initial visit (Baseline) and the Time Two and Three visits for **34** families (i.e. 45% of the original sample). Reasons for attrition at each of the follow-up points are shown in table 5. Across the whole of the study period, data were most often not provided at follow-up visits because families' needs had been met and the service's support was no longer required (see first row, table 5).

Table 5. Reasons for absence of data at follow-up visits

Reason for no data at follow-up points	Attrition from Time 1 to Time 2		Attrition from Time 2 to Time 3		Total Attrition Time 1 to Time 3	
	%*	(n)	%*	(n)	%*	(n)
Family no longer needed support – needs met	12	(9)	16	(12)	28	(21)
Family received visits but no data collected at follow-up	8	(6)	5	(4)	13	(10)
Family didn't engage/could not be contacted	5	(4)	0	(0)	5	(4)
Family moved away	4	(3)	1	(1)	5	(4)
Family referred on	4	(3)	0	(0)	4	(3)
Total attrition	33	(25)	22	(17)	55	(42)

* Per cent of original sample, N76. Figures may not add up to 100 due to rounding.

Demographic comparison of the families providing data at all three time points and those who provided data at the initial assessment but not at the latter two points showed there were no statistically significant differences in age, marital status, ethnicity, numbers of children, or employment status. Thus, although the sample by Time Three was small, we have no reason to believe that families remaining in the study to Time Three were unrepresentative of Home-Start parents more widely in terms of their social and demographic characteristics.

9.4.2. Support received

In addition to collecting outcome data, Co-ordinators recorded information about the number of volunteer visits and the duration of support from Home-Start. Volunteer visits occurred on average at intervals of between one and two weeks, and the number of volunteer visits to families provided between Baseline and Time Three ranged from 2 to 29, with an average of 13.4 (s.d.=6.2, n=60). The duration of support provided by Home-Start is shown in table 6. It shows that around half of families (47%) were no longer receiving support at the end of the study period, and 46% continued to receive support beyond Time Three. If we exclude the families with missing data, the proportion of parents receiving support for up to six months is 37%, which is broadly comparable to Home-Start's own figure of 45% for the service nationally. This suggests that the families taking part in the pilot study were fairly typical of all families that Home-Start supports, in terms of duration of help provided.

Table 6. Duration of family support from Home-Start

Duration of support	%	n
Up to 1 month	4	3
1 month to 3 months	10	8
3 months to 6 months	20	15
6 months to end of study period (8 months)	13	10
Support still continuing at end of study period	46	35
Not recorded	7	5
Total	100	76

In section 9.5 we report on the changes in scores on the three measures assessed at the initial visit and Time Two for the 51 families with data at these time points. In section 9.6 we report on changes in scores on the measures for the 34 families with data from the Baseline to Time Three.

9.5. Changes from Baseline to Time Two: 51 parents

We compared scores between Baseline and Time Two on all three measures for data provided by the 51 families with data at these time points. For each measure, an increase in score was indicative of improvement. There were missing values for the *PSAM* measure and the *Enjoyment of Parenting* measure for one family, resulting in a sample size of 50 for the analysis of these two measures.

9.5.1 Changes in *Parenting Self Agency Measure* scores from Baseline to Time Two

When we compared parents' *PSAM* scores measured at first visit with those measured at Time Two, we found a small increase in scores towards greater self-agency. The average *PSAM* score at initial visit for this group of parents was at 18.7 (s.d.=4) compared with 19.6 (s.d.=3.8) at Time Two (n=50, 1 missing). This result was statistically significant ($p=.02$, Wilcoxon test⁸).

9.5.2 Changes in *Enjoyment of Parenting* scores from Baseline to Time Two

We also compared parents' responses to the item concerning 'I enjoy being a parent' as assessed at initial visit and Time Two. The average scores for this group changed relatively little from Baseline to follow-up visit at Time Two, with averages of 4.4 (s.d. = .8) and 4.5 (s.d. = .7) respectively (n=50, 1 missing), and this change was not statistically significant.

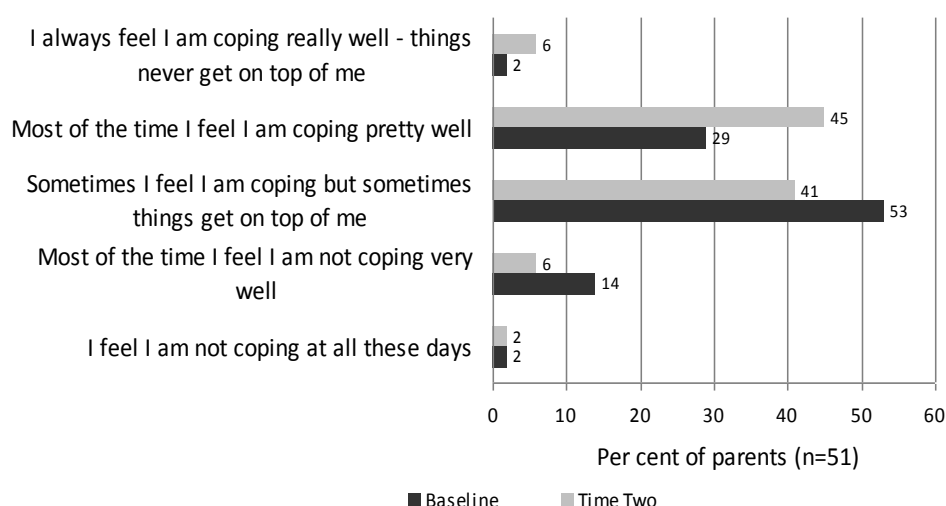
⁸ The Wilcoxon matched-pairs signed-ranks test compares the differences between the scores of individual respondents on their first completion of measures and their subsequent completion of measures in terms of their ranked size and direction (i.e. improvement or deterioration). It is considered to be a more cautious, more conservative test suitable for use with non-parametric data (those that are not normally distributed, as in this study).

9.5.3 Changes in the *Parent Coping Scale* scores from Baseline to Time Two

Finally, we compared scores on the *PCS* at the time of the initial visit with that of the first follow-up visit. The proportion of parents endorsing each of the statements for each visit is shown in figure 3. It can be seen that the proportion of parents rating themselves as coping 'most of the time' rose from 29% at the initial visit to 45% at the Time Two visit. The average score on this item changed from 3.2 (s.d. = .8) at Baseline to 3.5 (s.d. = .8) at Time Two for this group. Thus parents were rating their coping more positively at Time Two compared to the initial visit, and this change in parents' coping ratings was statistically significant at the 99% level of confidence ($p=.002$, $n=51$, Wilcoxon test).

We found that 19 of the 51 (37%) parents reported improved coping, 28 remained the same (55%), and 4 (8%) reported worse coping between Baseline and Time Two. When parents who reported improved coping scores were compared to those who had not changed or got worse, there were no differences in demographic characteristics such as age, marital and employment status, number of children, and ethnicity.

Figure 3 Changes on the *Parent Coping Scale* at Baseline and Time Two



9.6 Changes from Baseline to Time Two and Time Three: 34 parents

9.6.1 Responses at each wave of data collection

There were 34 families who provided data at all three time points in the study (i.e. Baseline, Time Two and Time Three). One family had missing values for the *PSAM* and *Enjoyment of Parenting*, resulting in a sample size of 33 for the analysis of these two measures. We compared scores for these parents on each of the measures at each of the three time points, and the means, standard deviations (s.d.) and results for tests of statistical significance for each measure are shown in table 7.

Table 7. Means and standard deviations for scores at Baseline, Time 2 and Time 3

Measure	Time 1 (Baseline) Mean (s.d)	Time 2 Mean (s.d)	Time 3 Mean (s.d)	Significance
<i>PSAM</i> N=33	19.2 (3.5)	19.5 (3.8)	20.4 (4.2)	Non-significant
<i>Enjoyment of parenting</i> N=33	4.5 (.7)	4.5 (.8)	4.6 (.6)	Non-significant
<i>Parent Coping Scale</i> N=34	3.1 (.8)	3.4 (.7)	3.7 (.9)	p<.001

Results of the statistical analysis of the changes in scores are described separately for each measure below.

9.6.2 Changes on *Parenting Self Agency Measure* scores across the three time points

Although there was a substantive improvement in this group of parents' rating of their self agency at each follow-up visit, unlike the changes in the larger group of parents completing only Baseline and Time Two questionnaires, these changes were not of sufficient magnitude to reach statistical significance. This was the case whether we compared their scores from Time One to Time Two, or from Time Two to Time Three, or from Time One to Time Three.

9.6.3 Changes in *Enjoyment of Being a Parent* scores across the three time points

As can be seen from table 7, parents' scores on their enjoyment of being a parent improved by a very small margin between Baseline and Time 3, and this change did not reach statistical significance. Changes in scores on this measure were not statistically significant when compared across any of the time points.

9.6.4 Changes on the *Parent Coping Scale* scores across the three time points

Amongst the group of parents providing data at all three time points, ratings of coping with being parent increased from Baseline to Time Two, and increased again at Time Three. The changes at each time point were statistically significant (Time One to Time Two $p<.01$; Time Two to Time Three $p<.05$; Time One to Time Three $p<.001$; Wilcoxon), indicating a continuous improvement in coping over the study period.

Between Baseline and Time Three, we found that 19 of the 34 (56%) parents reported improved coping, 13 (38%) remained the same and 2 (6%) reported that coping had deteriorated. When parents who had reported improved coping scores were compared to those who had not changed or had got worse over the time period, there were no differences in demographic characteristics such as age, marital and employment status, number of children, and ethnicity.

9.7 Feedback on process issues from Co-ordinators

9.7.1 Why feedback is important

It is striking that few evaluation studies collect process data relating to the methods of data collection. One of the original aims of this project was to develop a measure that could be administered by Home-Start staff or volunteers during the course of their normal work. In the pilot testing phase, Home-Start Co-ordinators took on the job of administering the questionnaire to parents at each of the three time points. Their responses to an open-ended 'feedback' questionnaire about process issues allowed us to explore the logistics and practicalities of introducing quantitative assessment measures into a service like Home-Start. They throw light on contextual factors potentially affecting the quality of the data that had been gathered. They also help to illuminate some of the psychometric properties of the three alternative measures that could influence their suitability to be used as an overarching measure of impact. The results below illustrate how valuable such feedback can be, and how it materially contributes to debates about the interpretation of findings.

We asked for Co-ordinators' feedback at the point when all Baseline, and some Time Two data, had been collected. The schedule used can be viewed in Appendix 4, but in brief, we asked for their perspectives on the following issues:

- Ease of explanation to parents
- Parents' understanding of the purpose of the data collection
- The acceptability of the process to parents
- Any specific barriers to completion
- Changes in parents' attitudes to participation over time
- Parents' understanding of the question wordings
- Whether any social desirability response bias was apparent (faking good or bad)
- Parents' attention and engagement with the questions
- Specific difficulties in generalising answers about different children in the family
- Co-ordinators' own views on if and how the process could be improved

Thirteen Co-ordinators who ran 12 schemes completed questionnaires (with two Co-ordinators each contributing from one large scheme that they ran jointly). All of their responses were thematically labelled and charted in relation to a number of process issues set out in each section below. We have used verbatim quotes from their written responses to illustrate the points made.

9.7.2 Introducing the evaluation to parents

Questionnaire responses indicated that most Co-ordinators felt it was generally easy to introduce the project to parents. However, more than half commented that although they could explain the evaluation to parents easily, it either interrupted the usual flow of the initial assessment visit or overly extended the visit.

'I found that it was easier to introduce and explain to some families than others. It really affected the natural flow of my initial visits, which I always try to make as

informal and unobtrusive as possible. For families wary of bureaucracy it presented a problem as it appeared to put them on the defensive, for other families it was time-consuming when they were trying to manage difficult children, etc.'

At initial visits, Co-ordinators explain the service to parents as well as assess their needs, so having to additionally explain the evaluation measures made the visiting time feel pressured. Sometimes it was difficult for Co-ordinators to hold a detailed discussion of the evaluation aims when parents had also to focus on managing the family, for example, keeping toddlers entertained or feeding a baby.

'The difficulties were purely practical – time and peace. The children within the family ran out of patience waiting for us to finish the adult conversation and I had no way of making this easier to manage.'

Only one co-ordinator commented that it fitted 'well' within the routine assessment visits. Generally, there was a strong sense that Co-ordinators did not especially welcome the burden of this extra task, even though the questionnaire was comparatively short.

9.7.3. Parents' understanding of the purpose behind the evaluation

We did not collect feedback directly from parents and are only able to report here on Co-ordinators' impressions of parents' understanding of the purpose of the measures. Reassuringly, many Co-ordinators reported that parents found it straightforward to understand the purpose of the measures, of which this response is a typical example:

'All families appeared to understand the purpose of the project.'

Nevertheless, a substantial proportion (around a quarter) felt that some parents didn't appear to understand the purpose, especially when they had already given verbal information as part of Home-Start's usual initial assessment.

'It wasn't easy for some of them to understand why we needed extra information, i.e. Home-Start initial assessment forms, then another form.'

Lack of clarity about the duplicative nature of the questions was understandable, however, and perhaps unavoidable, given that around the same time that some of the data collection took place, Home-Start was also introducing new assessment forms using questions about 'coping' via the MESH system into local Home-Starts in Northern Ireland (see Section 8).

9.7.4 Parents' concerns about the evaluation

As many as half of the Co-ordinators reported that at initial visits, families appeared concerned about judgements being made about their parenting, and that parents were sometimes anxious about the consequences of engagement with outside family support services. This is not an uncommon reaction upon initial contact with support services (see Ghate and Hazel 2002 for discussion about negative feelings towards helping agencies in a large national study), but Co-ordinators noted the possibility that in some cases, these

more general anxieties may have affected the way that parents completed the questionnaire.

'My families with social services involvement in particular could not be convinced of the benign nature of the (evaluation) project and I feel certain that this impacted on their responses. Nobody refused to participate, but their responses were measured.'

'It was okay, but I had to stress that it was all anonymous and their names were not included in the project.'

Another group of parents may have had difficulties completing the questionnaire measures because English was not their first language, or else they had literacy problems.

'Some mothers were a bit wary. Some mums asked me to fill it in for them, some due to literacy problems, and some due to feeding or amusing a tired child.'

9.7.5 Parents' understanding of the wording, and engagement with the measures

Although some parents had literacy difficulties and English was not their first language in some cases, the majority of parents appeared to understand the wording of the outcomes measures. *'No problem'* was the most frequent response from Co-ordinators to questions about parents' difficulties in the understanding of item wording.

However, there were several comments from Co-ordinators specifically about the apparent differences they observed between parents' ability to respond to the *PSAM*, the *Enjoyment of Parenting Measure*, and the *Parent Coping Scale (PCS)*. It seemed that around a quarter of Co-ordinators thought that parents found the *Parent Coping Scale* easier to respond too.

'The first questionnaire [the PSAM] seemed to cause confusion.'

'Because most of the questions were very similar, with little variation, it caused some confusion [on the PSAM].'

'The statements on the back of the questionnaire [ie, the PCS] were more accurate in assessing how that parent felt in general.'

'No difficulty was mentioned but it was simpler to complete the second set [i.e. the PCS] as only one statement had to be circled.'

Finally, two items in the *PSAM* require an answer related to parenting of a specific child in the family, who would have to be pre-identified to be the focus of questioning. Home-Start's model of working does not identify one particular child in a family to be a focus of support, so we adapted and generalised the wording to refer to 'my children' rather than my child'⁹.

⁹ For example, "I can solve most problems between *my child* and me", which we instructed parents to answer as: "I can solve most problems between *my children* and me". We are grateful to Larry Dumka for his permission to modify the *PSAM* in this way.

However, eight of the thirteen Co-ordinators commented that our adaptation was not successful, and parents found it difficult to respond to these questions:

'I think parents found that very difficult and tended to answer it for all children, but then discussed how it was different for each child.'

Engagement with the measures appeared to be good. There were two reported cases in which parents completed the questionnaire very rapidly without seeming to reflect on the items. Otherwise, most Co-ordinators reported that parents were completing the questionnaires in a careful and considered way.

9.7.6 Response bias: Social desirability response bias ('faking good' and 'faking bad'), and response-shift bias

Finally, a major concern for us was whether parents would feel able to answer the questions completely frankly, and would give Baseline responses that would not be subject to downwards re-appraisal once the service commenced delivery as an artefact of measurement rather than as a reflection of the quality of the service. As discussed earlier in Section 6.1, social desirability response bias ('faking good' and 'faking bad') is a recognised threat to methodological validity in evaluation research that requires people to answer questions about sensitive or personal matters. This can occur even when the questions are presented by an independent researcher, but is especially likely if questions are presented by someone known to the respondent or someone in a position to influence the respondent's situation. Response-shift bias might also have been an issue for this study, if contact with volunteers after the initial visit made parents more aware of ways in which they were not parenting optimally, and led them to reappraise their initial self-judgements in a negative direction, in ways unrelated to the service's impact on their actual coping. We discuss this in some detail below given its importance for the study, drawing on the insights provided by Co-ordinators who were asked to give an assessment of the extent of this problem, if present, using their own knowledge of the respondents' personal circumstances.

We anticipated that in the pilot study, these response biases might operate in one of three ways: (1) that parents might under-report difficulties ('fake good') in order to present themselves in a better light to Home-Start staff. This might especially be a feature of early contacts, before a trusting relationship had been established, and so could skew the Baseline results. (2) that parents might over-report difficulties ('fake bad'), perhaps in order to be assured of getting support they wanted, or ensuring that valued support was not withdrawn. This might also skew results, both at Baseline and potentially at later time points. (3) If response-shift were occurring, parents might report worse coping, self-efficacy or enjoyment at Time Two than at Baseline, (although we might expect to see rates rise again by Time Three if the service was providing effective help).

Co-ordinators' feedback confirmed that the first two of these biases appeared to be operating, to a different extent and in different ways in the data collection. For example,

around three-quarters of Co-ordinators commented that parents sometimes appeared to be 'faking good', i.e. parents were rating themselves more positively on the questionnaire items than the Co-ordinator had anticipated from prior conversations. For example, one Co-ordinator noted that a parent whose several children had been taken into care nevertheless rated herself at the most positive end of each scale on all the measures at the initial visit. Some Co-ordinators felt that parents found it difficult to acknowledge 'on paper' that they were not coping well due to concerns about how the written information would be used. Others felt that parents were simply trying hard to be positive and were reluctant to admit to not coping. For some this meant '*sticking to the middle ground*' as one Co-ordinator put it, and rating themselves at least on the middle of scales rather than at the less positive end. One Co-ordinator also felt that in a small number of cases parents who had been referred by other agencies were ambivalent about whether they wanted Home-Start's support, and rather than say this directly to the Co-ordinator or the person who had referred them, the parent used the questionnaire as a means of demonstrating that the extra support was not needed (ie, by rating themselves very positively).

Co-ordinators felt that 'faking good' was happening frequently among the families: '*In almost every case*', '*...about 60% of the time*', '*Regularly*':

'I feel parents found it hard to say they are not coping or give a very low score, even though the conversation indicated that this was the case at times.'

'Because most of our families viewed this as a judgement they tended to rate themselves pretty high even though it had no bearing on the actual reason they needed Home-Start support....there was no room for us to 'improve' their situation based on this as a measurement scale.'

'I think parents like to think things are ok or are getting better and will err on the side of being more positive. However, I also think they answered as honestly as possible. As the relationship with the family developed I feel that parents felt more open and relaxed.'

'One parent 'confessed' to 'faking good' on the first questionnaire only. She 'confessed' this when I gave her the second one.'

'Faking bad' seemed, on the other hand, to be far less common, though not completely absent. Co-ordinators reported that faking bad appeared to happen with two families. One of these cases involved a parent who in the Co-ordinator's judgement was a particularly heavy consumer of support. This parent might have been anxious about what would happen when the service came to an end.

Response-shift reappraisal, by contrast, did not appear to be operating at all. All measures showed on average either a degree of improvement or no change over time, rather than deterioration.

Thus, overall, Co-ordinators' feedback confirmed what inspection of the data at Baseline had previously suggested; ie, that a degree of response bias was present, especially for the first two measures we tested, and mainly in relation to social desirability response bias operating in a positive direction.

9.7.7 Co-ordinators' reflections on the evaluation

Finally, we asked Co-ordinators for their thoughts on their experiences of the evaluation as a whole. Among the 13 respondents, two felt so negative that they would not want to carry out the data collection again. For one person this was because they felt the inclusion of questionnaires in addition to the information that is routinely collected at visits added too great a burden, along with the time involved in extra administration and paper work such as photocopying. The second Co-ordinator felt that having to complete questionnaires with families adversely affected her approach to working with them.

'I feel that the families I have done it with I have not got to know as well as usual due to the more general process and what was seen as a question filling exercise.'

Another Co-ordinator also expressed concern about the way in which completing forms with a family might have impacted upon them, and as a result was selective when approaching families to take part in the project:

'There were some families I did not introduce the project to because their needs were so great and I felt that by introducing another set of forms may put them off accepting our service.'

As can be seen, the process feedback from Co-ordinators was insightful and helpful in clarifying some methodological issues. They were frank about their reservations and anxieties about using a structured measure in the context of the sensitive work of building and maintaining a trusting relationship with vulnerable families. Some would have preferred not to have been asked to participate. However, on balance, Co-ordinators were accepting of the value of the exercise, and keen to help refine the process of administration. When we also asked Co-ordinators for their thoughts on how the evaluation might have been improved, they suggested the following:

- Consult with volunteers (as well as Co-ordinators) about the relevance of a outcome construct
- Introduce the study after initial assessment visits but prior to the first volunteer visit
- Allow a longer interval between follow-up visits
- Allow a longer overall follow-up period

Section 10 Conclusions, implications and recommendations

10.1 The overall approach: broadly successful

We began this report by outlining the challenges of developing robust and fit-for-purpose evaluation measures for family support services like Home-Start that operate on a tight budget, are open-access, responsive and user-led and do not adhere to tightly specified treatment models underpinned by formally articulated theories of change. This kind of service probably describes the majority of family support provision in Europe today, notwithstanding the rise of the ‘evidence-based programme’ movement. We noted that this kind of service is especially difficult to evaluate satisfactorily using quantitative measurement techniques.

Broadly, our focus on *overarching impact* rather than more fine-grained, *domain-specific outcomes* as a response to this problem appears to have been fruitful. The methods used to develop the constructs to be tested (self-efficacy, self-agency and coping), appear to have satisfied the criterion of identifying a set of suitably global, overarching constructs that broadly reflect what Home-Start staff themselves regard as an essential, core purpose of the service: that is, to increase parents’ sense of being able to successfully contend with and manage the daily challenges of parenting.

The methods used to develop and test the alternative measures also appear to have been appropriate to the task and resulted in the selection and field testing of three plausible alternative measures based around pre-existing (and therefore pre-tested) single and multiple item scales. Although all measures tested were confirmed as conceptually related, and all were broadly acceptable to parents, they nevertheless demonstrated different psychometric properties. The parallel testing of three alternatives in the field allowed us to compare and contrast these, and select one measure as superior (see below, Section 10.2.2).

Overall, the results of the pilot were encouraging. Due in no small part to the efforts of all the Home-Start staff who helped to organise the field testing and collect the data, we successfully developed and administered three alternative measures of overall impact and gathered data at three time points over a 23 week follow-up period in 12 local Home-Starts in Northern Ireland. Using these data, the study was successful in identifying a simple, single overarching measure of impact that was responsive enough to pick up change in a medium-sized sample of Home-Start service users even taking account of attrition over time. Moreover, we believe the approach we developed will generalise well for use in other family support contexts, with some adjustments for specific differences.

10.2 Practical implications and limitations

10.2.1 Attrition between waves of data collection reduces the reliability of the conclusions

A number of practical and methodological issues were highlighted during the field testing and analysis stages of the project. First, it is important to note the rate of attrition over the

total study period was 55%, including 33% from Baseline to Time Two. Combined with the data on reasons for absence of data at follow-up points (see table 5, section 9.4.1), this means we need to be somewhat cautious about the findings. Although this rate is by no means outside the acceptable range in evaluation research, the 33% of parents who did not complete a questionnaire at Time Two and the further 22% who did not complete a questionnaire at Time Three may, possibly, have answered very differently the parents who did. From our data we can say that those who remained in the study to Time Three did not differ demographically from those who dropped out; but beyond this, there are unknowns. Parents who stayed in the sample might have been experiencing more positive results from Home-Start's help than those who dropped out earlier. Or, they may have had greater needs for support that we could not measure at the outset, and therefore may have been more likely to benefit. A larger sample might have generated different results. It might also be that a longer study, tracking changes in coping amongst those who leave early as well as those who stay, would produce different results to those reported here. To increase confidence in the results, further testing would be desirable to understand more about what underlies attrition between the time points.

Secondly, we cannot know whether the impact we recorded endures beyond the medium term (i.e., approximately eleven and twenty three weeks post commencement). Ideally, data collection should also continue beyond the sixth month milestone that usually marks the end of Home-Start's involvement, to capture more long term impact.

10.2.2 Measurement issues: one measure worked better than others

All the measures we selected for testing were assessed to be theoretically valid measures of overall impact of Home Start services, and all had the advantage of being extensively pre-tested in previous studies. At least one (*Parenting Self Agency Measure, PSAM*) had good detail on psychometric properties, and all had been tested on general population samples and thus 'normed' data were available against which to compare our results. However, not all appear to have worked equally well in the context of this feasibility study.

First, we noted earlier that two of the three measures we tested appeared to be subject to 'ceiling effects', such that the majority of parents were placing themselves at the upper (most positive) end of the scale at Baseline. Although ceiling or floor effects may reflect genuine lack of variability in a population, social desirability response bias can also produce effects of this kind. It is self-evident that any measure selected should be subject to the minimum possible social desirability response bias, not least because measures with strong ceiling (or floor) effects are not useful for the evaluation of change over time. On the *PSAM*, on four of the five items between one half and three quarters of respondents placed themselves at the upper end of the scale at Baseline. On the *Enjoyment of Parenting* measure, four fifths (82%) of parents did this. The single-item *Parent Coping Scale* by contrast seemed more robust. Baseline results on this measure were more dispersed, with only one third (32%) placing themselves at the most positive end of the scale at Baseline, and nearly one in six respondents admitting to coping poorly (not coping well/not coping at all). Further comfort comes from comparison of the scores on the *PCS* at Baseline with those in a large, nationally representative sample of parents with comparable demographic make-up but who were not actively seeking help (Ghate and Hazel 2002). This suggests - plausibly - that parents using Home-Start report relatively more difficulties.

Second, two of the measures tested (*PSAM* and the *Parent Coping Scale*) showed sensitivity to change and picked up significant changes between the first two waves of data collection. However, when we assessed sensitivity to change across the longer time frame of all three waves of data collection (from Baseline to Time Three: just under six months' duration), only the *Parent Coping Scale* recorded significant results.

Third, not all the questions were equally easy to administer and respond to. Co-ordinators reported that some parents struggled to understand and respond to the *PSAM* questions, and some questions (for example, those that concerned parent-child relationships) did not generalise well and were difficult to answer in a way that was not domain-specific or person-specific.

Although we can only speculate on the reasons for these results, several possible explanations can be offered:

The *Parenting Self Agency Measure* is worded in a relatively formal way. All the constituent statements are worded positively, which may have the effect of somewhat 'leading' the respondent. It also appeared first in the sequence of three measures, which may have affected how parents responded. Co-ordinators reported that parents struggled to answer some questions that were person-specific in a more generalised way.

Enjoyment of Parenting is a measure that produces substantial ceiling effects in studies in which it has been used. For example, in the large scale study of parenting in poor environments, rates of reporting enjoying being a parent 'all or most of the time' ranged from 92% to 88% depending on which child was the subject of the question (Ghate and Hazel, unpublished data). It may be the case that enjoyment of parenting operates independently of objective circumstances, and that most parents do, indeed, enjoy parenting all or most of the time even when they are stressed. Or, it may be that the statement is subject to such strong social norms that it is difficult to answer except in the affirmative. Whatever the explanation, our results suggested that this question is not useful as a measure of change over time, except perhaps - and this may be an important exception - for a small minority of extremely stressed and likely very high risk parents who admit to low enjoyment of parenting at the commencement of service delivery.

The *Parent Coping Scale* appeared last in the sequence, which may have affected how parents responded to the measure (they may have felt more familiar with the process of answering questions by this point, for example). However, it is also prefaced by an informally worded 'permission-giving' statement ("most people find that parenting has its ups and downs") and uses colloquial plain English throughout. The constituent statements are worded in a mix of negative and positive, thus allowing a range of responses. The statements do not relate to specific people or specific domains of parenting but invite parents to talk about overarching perceptions of parenthood as a broad role. Finally, the separate test-retest procedures undertaken by Home-Start indicate that the measure's reliability and stability over time points is excellent.

Thus we concluded that the *Parent Coping Scale* appeared to be the more robust and more useful measure for the purpose of measuring overarching impact. Consistent with our

attempts to measure overarching *impact* rather than more fine-grained domain-specific or person-specific *outcomes*, the *Parent Coping Scale* was focused on parents' perceptions of the parenting role in a wide sense rather than their parenting of a specific, named child. Its ready generalisation to all parenting circumstances irrespective of age or number of children permits the measure to be used across a wider variety of family situations and service responses than a measure that isolates and focuses on a single index child.

10.2.3 Data collection burden on Co-ordinators and lack of independence in data collection and sampling may subvert the reliability of the findings

The practicalities of introducing quantitative measurement to a service such as Home-Start may bear further consideration. Despite our strenuous efforts to make the data collection process as simple and minimal as possible, some Co-ordinators were clearly not entirely positive about the data collection enterprise. Asking staff to take charge of this process increases the burden upon them, and potentially influences both the way the services are received by parents at the outset, and the honesty with which parents respond to the evaluation question. In addition, asking service staff to administer the questionnaire also blurs the important distinction between questions for the purposes of assessment of need, and questions for the purpose of evaluation of service effectiveness, especially when there is overlap between the types of questions asked. It was perhaps not surprising, therefore, that there were substantial indications in feedback of 'faking good' in relation to some questions, and although some degree of social desirability response bias may be unavoidable when asking personal questions, this effect is likely to be heightened when the interviewer is personally known to the interviewee.

In addition, an issue about the representativeness of the pilot sample emerged via Co-ordinators' feedback. The sampling protocol designed for the feasibility study instructed Co-ordinators to invite **all** parents new to Home Start from October 2010 to participate. However, in feedback, some Co-ordinators reported that they had made informal decisions during the pilot study about whether or not to invite parents to take part. Parents who were thought to have more fragile or ambivalent attitudes towards Home-Start were not invited, on grounds that being asked to complete the questionnaire could act as a further disincentive. Although this is an understandable decision when viewed from the service perspective, from an evaluation perspective it is problematic. Informal selection into or out of the sample effectively subverts the principle of a representative sample and will tend to bias the eventual results towards the positive by selecting out the more 'difficult' cases.

10.2.4 Absence of a counterfactual limits the robustness of the conclusions

Finally, a major limitation to the methodology of the pilot study, which constrains our ability to interpret the results, was that we had no 'counterfactual' against which to compare the results. In other words, the study had no comparison group of parents not using Home-Start against which changes over time can be compared. The measure of 'distance-travelled' for the sample is valid in its own, internal, terms, but we cannot know whether Home-Start services caused these changes. Being cautious about our use of terminology, the best we can say is that use of Home-Start was *associated with* change over time, but we do not know that it *caused* it.

10.3 Recommendations

The results of the project led us to make the following recommendations, which incorporate some modifications to the process we trialled:

10.3.1 Recommended measure of overarching impact: the *Parent Coping Scale*

For future measurement of overarching, high level impact of Home-Start's services we recommend the use of the *Parent Coping Scale (PCS)*, which, of the three measures tested, appeared to achieve the best balance between practical, theoretical and scientific considerations. The *PCS* provides a global measure of 'coping with being a parent', a theoretically and empirically robust construct that is conceptually related to parenting self-efficacy and shown in empirical studies to be related to parenting practices, quality of parenting and a host of risk and protective factors for parenting and child development outcomes. Improving parents' sense of coping with parenting stress is recognised by Home-Start's stakeholders as a valid indicator of the intentions of the service, and 'coping' has a plain English meaning readily understandable to parents of all social and educational backgrounds. It has excellent stability over time points. Future development of the measure should however ideally test its behaviour in a comparison (non-service group).

Box 1 *Properties of the Parent Coping Scale*

The *Parent Coping Scale* demonstrated the following practical properties:

- Prefaced by a permission-giving, normalising statement worded in plain and colloquial British English that recognises variations in the daily 'lived experience' of parenting
- High acceptability to parents
- High saliency and general rather than domain or person-specific: widely applicable to parents of both sexes and in diverse situations, focusing on the broad parenting role, rather than specific parenting practices or relationships with specific children
- Uses a simple five-point response scale, formulated as statements
- Extremely brief to administer, either self-completed or interviewer-administered
- Easy to analyse
- Precursor measure normed in large-scale population surveys, providing comparable data for use in family support research

It also showed the following measurement and psychometric properties:

- Good face and content validity
- Good construct validity
- Psychometric testing indicates excellent stability and reliability
- High sensitivity to detect changes over time, in tests over a six month follow-up, even in a small sample
- Adequate dispersion and reduced social desirability response bias

Full details of the *Parent Coping Scale (PCS)*, its psychometric properties, and how to use it, are available on the web at:

http://www.cevi.org.uk/docs/Parent_Coping_Scale.pdf

10.3.2 Recommended method of data collection

Given the brevity and simplicity of the *PCS*, we also recommend that future use should explore the use of telephone rather than face-to-face administration methods. This would remove the need for local Co-ordinators or other Home-Start personnel to administer the measure, reducing both the burden on the service and the potential for social desirability response bias that arises when a person known to the respondent is associated with administration of the measure. Co-ordinators would still need to seek permission for on-going telephone contact, but data could then be collected centrally, by an independent researcher or telephone research unit. This would allow for speedy and cost-efficient data collection, more accurate timing of follow-ups, and also permit collection of data from parents who have completely ceased to use the service. Longer follow-up periods could also be employed, increasing the size and utility of the data-set for monitoring the longer-term impact of the service. All of these factors, if implemented, may yield new insights into the workings of the *PCS* and will contribute to the further refinement of the methods described in this study.

10.3.3 Recommended future testing to include a counterfactual

The absence of a counterfactual or comparison group means that we do not know how the measures might have performed in a group of parents who were not receiving Home-Start's services. It is possible that these measures would pick up the same degrees of change (or lack of it) irrespective of whether parents were receiving a service, or irrespective of whether that service was Home-Start or something else. Ideally, if this work proceeds to the next stage and the use of these measures is rolled out more widely, efforts should be made to identify a suitable counterfactual that can be used for further test purposes. This could perhaps be arranged as part of other studies that are taking place for other purpose, for example, by incorporating the selected measure within another study. Or, at relatively low cost, and especially if telephone interviewing methods are used, a counterfactual test could be set-up specifically for this purpose.

10.4 Overarching conclusions

In conclusion, the project has satisfactorily demonstrated that it is possible to develop and use a simple, low-cost overarching measure of impact for an open-access, user-led and community-based family support service delivered by volunteers across the UK. We were able to develop and implement a measure that was compatible with the service's broad intentions, acceptable to parents and staff, and satisfactorily robust as a measure of change over time. One measure in particular, the *Parent Coping Scale*, showed greatest sensitivity to change and holds particular promise as a simple, low-cost measure that could be used by Home-Start and potentially also by other community-based services with similar high-level goals. The project also has much wider applicability, in having developed a workable methodology for the development of other measures, in the case of services that seek other, different kinds of overarching impact.

We make a final observation with relevance not just to this project but to the field of evaluation of family support as a whole. We started the report by observing that qualitative and quantitative evaluation results often stand in puzzling contradiction to one another.

The attempt to capture the broad, 'big picture' impact of a service, described in this report, may have more in common with qualitative research than other quantitative measurement approaches. In avoiding the pitfalls of trying to match fine-grained, domain-specific outcome measures to the diffuse objectives and fluid operations of many non-programmatic family support services, impact measurement, like qualitative enquiry, also relies on the synthesis of multiple contributory factors into overarching summative judgments. In this respect, well-constructed measures of impact may help towards the bridging of the disconnection that often emerges when qualitative and quantitative research are employed together to assess the results of family support services.

References

References

- Abidin, R. R. (1983) *Parenting Stress Index Manual*. Charlottesville, VA: Pediatric Psychology Press.
- Abidin, R. R. (1995) *Parenting Stress Index. Third Edition: Professional Manual*. Odessa, FL: Psychological Assessment Resources, Inc.
- Altmaier, E.M. (1995) Linking stress experiences with coping resources and responses: Comment on Heppner et al. (1995), Bowman and Stern (1995), Catanzo et al. (1995), and Long and Schutz (1995). *Journal of Counseling Psychology*, 42, 3304-3306.
- Asscher, J., Deković, M., Prinzie, P., and Hermanns, J. (2008a) Assessing change in families following the Home-Start program: clinical significance and predictors of change. *Family Relations*, 57, 351-364
- Asscher, J., Hermanns, J., and Deković, M. (2008b) Effectiveness of the Home-Start parenting program: behavioural outcomes for parents and children. *Infant Mental Health Journal*, 29, 95-113.
- Asscher, J., Hermanns, J., Deković, M., and Reitz, E. (2007) Predicting the effectiveness of the Home-Start parenting support program. *Children & Youth Services Review*, 29, 247-264
- Aye-Maung N; Parfremment J and Tipping S (2008) *National Evaluation of On Track Phase Two: final report of the longitudinal cohort study* London: Department for Education and Skills
- Bandura, A. (1977) Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84, 191-215.
- Bandura, A. (1982) Self-efficacy mechanism in human agency. *American Psychologist*, 37, 122-147.
- Bandura, A. (1997) *Self-efficacy: the exercise of control*. New York: WH Freeman Publishing.
- Bandura, A. (2006) Guide for constructing self-efficacy scales. In F. Pajares and T. Urdan (Eds.) *Self-efficacy beliefs of adolescents*, 5, pp. 307-337. Greenwich, CT: Information Age Publishing.
- Barlow J, Davis H, McIntosh E, Kirkpatrick S, Peters R, Jarrett P and Stewart-Brown S, (2008) *The Oxfordshire Home Visiting Study: three year follow-up* <http://wrap.warwick.ac.uk/43244/> and <http://www.herc.ox.ac.uk/research/homevisitttrial> [accessed 23rd April 2012]
- Barnes, J., MacPherson, K. and Senior, R. (2006) The impact on parenting and the home environment of early support to mothers with new babies. *Journal of Children's Services*, 1, 4-20.
- Barrett, H. (2007) *Evaluating evaluations. Evaluating recent evaluations of Sure Start, Home-Start and Primary age learning study*. London: Family and Parenting Institute.
- Biehal N, Dixon, J, Parry E, Sinclair, I, Green J, Roberts C, Kay C, Rothwell J, Kapadia D and Roby A (2012) *The Care Placements Evaluation (CAPE) Evaluation of Multidimensional Treatment Foster Care for Adolescents (MTFSC-A)* <https://www.education.gov.uk/publications/standard/publicationDetail/Page1/DFE-RR194> [Accessed June 24th 2012]

Bowling, A. (2005) Just one question: if one question works why ask several? Editorial. *Journal of Epidemiology and Community Health*, 59, 342-345.

Bloomfield, L. And Kendall, S. (2007) Testing a parenting programme evaluation tool as a pre- and post-course measure of parenting self-efficacy. *Journal of Advanced Nursing*, 60, 487-493.

Brown, K. and Burrows, C. (1992) Working paper 26: How do we measure 'change' in utility measures in health status – or should we? Centre for Health Program Evaluation. [online] Available at:

<<http://www.buseco.monash.edu.au/centres/che/pubs/wp26.pdf>> [Accessed 23 July 2010]

Campbell D.T. and Stanley J.C. (1963) *Experimental and quasi-experimental designs for research* Chicago: Rand McNally

Centre for Study of Social Policy, (2003) Strengthening families through early care and education [online]

Available at: <http://www.cssp.org/uploadFiles/horton.pdf> [Accessed 25 June 2010]

Chen H.T. (1989) The conceptual framework of the theory-driven perspective. *Evaluation and Program Planning*, 12, 391-396.

Coleman, P. and Karraker, K. (1998) Self-efficacy and parenting quality: findings and future applications. *Developmental Review*, 18 (1) 47-85.

Cutrona, C. and Troutman, B. (1986) Social support, infant stress, and parenting self-efficacy: a mediational model of postpartum depression. *Child Development*, 57, 1507-1518.

Deković, M., Asscher, J., Hermanns, J., Reitz, E., Prinzie, P. and van den Akker, A. (2010) Tracing changes in families who participated in the Home-Start parenting program: parental sense of competence as mechanism of change. *Prevention Science*, 11, 263-274.

De Montigny, F. and Lacharité, C. (2005) Perceived parental efficacy. *Journal of Advanced Nursing*, 49, 387-396.

Diehr, P., Chen, L., Patrick, D., Feng, Z., and Yasui, Y. (2005) Reliability, effect size, and responsiveness of health status measures in the design of randomized and cluster-randomized trials. *Contemporary Clinical Trials*, 26, 45-58.

Dumka, L., Stoerzinger, H., Jackson, K., and Roosa, M. (1996) Examination of the cross-cultural and cross-language equivalence of the parenting self-agency measure. *Family Relations*, 45, 216-222.

Family Links Research Team at the University of Warwick and Swansea University (2011) *Promoting Health and Wellbeing with the Family links Nurturing Programme (FLNP) in South Wales: a randomised controlled trial and economic evaluation* [Accessed May 30th 2012]

<http://wales.gov.uk/statistics-and-research/promoting-health-wellbeing-family-links-nurturing-programme/?lang=en>

Fitzpatrick, R., Fletcher, A., Gore, S., Jones, D., Spiegelhalter, D. and Cox, D. (1992) Quality of life measures in health care. I. Applications and issues in assessment. *British Medical Journal*, 305, 1074 – 1077.

- Fixsen D.L., Blase K.A. Naoom S.F. and Wallace F (2009) 'Core implementation components' *Research on Social Work Practice* 19 (5) 531-540
- Foddy W (1993) *Constructing questions for interviews and questionnaires – theory and practice in social research* Cambridge : Cambridge University Press
- Folkman, S. and Lazarus, R. (1991) Coping and Emotion. In: Monat, A.M. and Lazarus, R.S. *Stress and coping*, pp. 207-227, New York: Columbia University Press.
- Frost, N., Johnson, L., Stein, M., and Wallis, L. (2000) Home-Start and the delivery of family support', *Children and Society*, 14, 328-342.
- Ghate, D. (2001) 'Community-based evaluations in the UK: scientific concerns and practical constraints.' *Children and Society* 15, 23-32 (invited contribution: special edition on Evaluating Community Initiatives for Children)
- Ghate, D. and Hazel, N. (2002) Parenting in poor environments: stress, coping and support. London: Jessica Kingsley.
- Ghate D., Asmussen K., Tian Y., and Hauari H. (2008): *Reducing risk and increasing resilience: how did On Track work?*
- Gilmore, L. & Cuskelly, M. (2008). Factor structure of the parenting sense of competence scale using a normative sample. *Child: Care, Health & Development*, 35, 48-55.
- Goodman, R. (1997) The strengths and difficulties questionnaire: a research note. *Journal of Child Psychology and Psychiatry*, 38, 581-586.
- Goodman, R. (1999) The extended version of the strengths and difficulties questionnaire as a guide to child psychiatric caseness and consequent burden. *Journal of Child Psychology and Psychiatry*, 40,791-801
- Hardiker P, Exton K and Barker M (1991) *Policies and Practices in Preventive Child Care* Aldershot: Avebury
- Harris, T. (2009) Listening better (and hearing) through psychotherapy research. *British Journal of Psychotherapy*, 25, 343-348.
- Harrison, M. (2003) *Hooray! Here comes Tuesday: The Home-Start story*. Bamaha Publishing.
- Home-Start (2010a) What we do. [online] Available at: http://www.home-start.org.uk/about/what_we_do [Accessed 10th January 2011].
- Home-Start (2010b) *A report of the statistics information collected from 334 schemes*. Available at: http://www.home-start.org.uk/about/statistics_0910 [Accessed 17th July 2011]
- Home-Start (2012) Statistics overview 2008/2009. [online] Available at: http://www.home-start.org.uk/about/Statistics_overview_08_09 [Accessed 10th January 2011].
- Howard G.S (1980) Response-shift bias: a problem in evaluating interventions with pre/post-test reports *Evaluation Review* 4 (1) 93-106 <http://erx.sagepub.com/content/4/1/93.abstract> [Accessed May 30th 2012]

Hyland, M.E. (2003) A brief guide to the selection of quality of life instrument. *Health and Quality of Life Outcomes*, 1, 24.

Izzo, C., Weiss, L., Shanahan, T., and Rodriguez-Brown, F. (2000) Parental self-efficacy and social support as predictors of parenting practices and children's socioemotional adjustment in Mexican immigrant families. *Journal of Prevention and Intervention in the Community*, 20, 197-214.

Johnston, C. and Mash, E.J. (1989) A measure of parenting satisfaction and efficacy. *Journal of Clinical Child Psychology*, 18, 167-75.

Jones, T.L. and Prinz, R.J. (2005) Potential roles of parental self-efficacy in parent and child adjustment: A review. *Clinical Psychology Review*, 25, 341-363.

Kendall, S and Bloomfield, L. (2005) Developing and validating a tool to measure parenting self-efficacy. *Journal of Advanced Nursing*, 51, 174-181.

Kirkaldy, A. and Crispin, A. (1999) Home-Start in Scotland: An Evaluation. Dunfermline: Home-Start UK.

Lamb T. (2005) The Retrospective Pretest: An Imperfect but Useful Tool *The Evaluation Exchange XI (2) Summer 2005*, 18-19 <http://www.hfrp.org/evaluation/the-evaluation-exchange/issue-archive/evaluation-methodology/the-retrospective-pretest-an-imperfect-but-useful-tool> [Accessed 1 Feb 2012].

Lazarus, R.S. (1993) Coping theory and research: Past, present and future. *Psychosomatic Medicine*, 55, 234-247.

Lazarus, R.S. and Folkman, S. (1984) *Stress, appraisal and coping*. New York: Springer.

Letarte, M.J, Normandeau, S. and Allard, J. (2010) Effectiveness of a parenting training programme 'Incredible Years' in a child protection service *Child Abuse and Neglect* 34 253-261

McAuley, C. (1999) *The family support outcomes study*. Northern Ireland Health and Social Services Board. Ballymena: Northern Ireland.

McAuley, C., Knapp, M., Beecham, J., McCurry, N. and Slead, M. (2004) *The outcomes and costs of Home-Start support for young families under stress*. York: Joseph Rowntree Foundation.

McHorney C.A., Ware J.E., Rogers W., Raczek A., and Lu, J.F.R. (1992) The validity and relative precision of MOS short- and long-form health status scales and Dartmouth COOP charts: results from the Medical Outcomes Study. *Medical Care*, 30 (Supplement 5), MS253-MS265.

McMahon, R.J., and Metzler, C.W. (1998) Selecting parenting measures for assessing family-based preventive interventions. In R.S. Ashery and E.B. Robertson (Eds), *Research meeting on drug abuse prevention through family interventions*. NIDA Research Monograph 177 (pp. 294-323), NIH Publication NO. 99-4135, Washington, DC: National Institute on Drug Abuse.

MacPherson, K., Barnes, J. Nichols, M. and Dixon, S. (2010) Volunteer support for mothers with new babies: perceptions of need and support received. *Children and Society*, 24, 175-187.

Mason, J.W. (1975) A historical view of the stress field. Part 1. *Journal of Human Stress*, 1, 6-12.

Maughan, D.R., Christiansen, E., Jenson, W.R., Olympia, D. and Clark, E. (2005) Behavioral parent training as a treatment for externalising behaviours and disruptive behaviour disorders: A meta-analysis. *School Psychology Review*, 34, 267-286.

Mitchell P F (2011) Evidence-based practice in real world services for young people with complex needs: new opportunities suggested by recent implementation science
Children and Youth Services Review 33 207-216

Moran, P., Gbate, D. and van der Merwe, A. (2004) *What Works in Parenting Support?: A Review of the International Evidence*. London: HMSO.

Pisterman, S., Firestone, P., McGrath, P., Goodman, J., Webster, I., Mallory, R. and Goffin, B. (1992) The effects of parent training on parenting stress and sense of competence. *Canadian Journal of Behavioural Science*, 24, 41-58.

Sanders, M.R. and Woolley, M.L. (2005). The relationship between maternal self-efficacy and parenting practices: implications for parent training. *Child: Care, Health and Development*, 31, 65-73.

Shinman, S. (2005) *Learning from Families. Policies and practices to combat social exclusion amongst families with young children in Europe*. Home-Start UK: Leicester.

Schorr L. (2003) *Determining 'what works' in social programs and social policies: towards a more inclusive knowledge-base* Children's Roundtable: the Brookings Institution
<http://www.lisbethschorr.org/doc/DeterminingWhatWorksInSocialProgramsandPolicies.Feb.2003.pdf> accessed May 30th 2012

Skinner, E., Edge, K., Altman, J., and Sherwood, H. (2003). Searching for the structure of coping: a review and critique of category systems for classifying ways of coping. *Psychological Bulletin*, 129, 216-269.

Stewart-Brown, S., Anthony, R., Wilson L., Winstanley S., Stallard, N., Snooks H. and Simkiss, D. (2011) Should randomised controlled trials be the 'gold standard' for research on preventive interventions for children? *Journal of Children's Services*, 6(4) 228-235

Stifter, C.A. and Bono, M.A. (1998) The effect of infant colic on maternal self-perceptions and mother-infant attachment. *Child: Care, Health and Development*, 24, 339-351.

Terwee C.B., Dekker F.W., Wiersinga W.M., Prummel M.F., and Bossuyt P.M.M. (2003) On assessing responsiveness of health-related quality of life instruments: guidelines for instrument development. *Quality Life Research*, 12, 349-362.

Teti, D.M. and Gelfand, D.M. (1991) Behavioural competence among mothers of infants in the first year: the meditational role of maternal self-efficacy. *Child Development*, 62, 918-929.

Utting, D. ,Monteiro H. and Gbate D (2007) *Interventions for children at risk of developing anti-social personality disorder* London: Policy Research Bureau and Cabinet Office

Vermeersch, D. A., Lambert, M. J., and Burlingame, G. M. (2000) Outcome questionnaire: item sensitivity to change. *Journal of Personality Assessment*, 74, 242-261.

Wanous, J. P., Reichers, A. E., and Hudy, M. J. (1997). Overall job satisfaction: how good are single-item measures? *Journal of Applied Psychology*, 82, 247-252.

Whetten D.A (1989) What constitutes a theoretical contribution? *Academy of Management Review* 14:4 490-495

Whittaker, K., and Cowley, S. (2006) Evaluating health visitor parenting support: validating outcome measures for parental self-efficacy. *Journal of Child Health Care* 10, 296-308.

Zeedyk, M. S., Werritty, I., and Riach, C. (2002) The PALS parenting support programme: lessons learned from the evaluation of processes and outcomes. *Children & Society*, 16, 318-333.

Appendices

Appendix 1. Parent questionnaire assessing outcomes

The statements below describe feelings and thoughts about being a parent. Please tell us how often you feel or think like the statement, **thinking about your current situation**.

A Please ✓ in the box that best describes you.

	1 Almost never or never	2 Once in a while	3 Sometimes	4 A lot of the time	5 Almost always or always
I feel sure of myself as a mother/father.					
I know I am doing a good job as a mother/father.					
I know things about being a mother/father that would be helpful to other parents.					
I can solve most problems between my child and me.					
When things are going badly between my child and me, I keep trying until things begin to change.					
I enjoy being a parent					

B Most people find being a parent has its ups and downs. Taking everything into account, which of these statements best describes how you are coping with being a parent these days?

Circle one number:

1. I feel I am not coping at all these days
2. Most of the time I feel I am not coping very well
3. Sometimes I feel I am coping but sometimes things get on top of me
4. Most of the time I feel I am coping pretty well
5. I always feel I am coping really well – things never or hardly ever get on top of me

Thank you for your help

For office use only:

Parent no:

Centre no:

Completed on visit:

1

2

3

Date of completion of this questionnaire:

Parent completed questionnaire independently ☐

Completed with support ☐

Appendix 2. Parental self-efficacy measures shortlisted but not selected

Tool to Measure Parental Self-Efficacy (TOPSE; Kendall and Bloomfield, 2005)

This measure was developed within the UK and tested with families attending parenting programmes. It contains 48 items in total, arranged in a number of subscales relating to different areas of parenting such as discipline, empathy and understanding. Although it has the advantage of having been developed within the UK, it was considered unsuitable within the context of the present report as most of the items refer to a 'target' child, and it is much lengthier than other measures that assess this construct.

Parental Sense of Competence Scale (PSOC; Gibaud-Wallston and Wandersman, 1978, cited in Johnston and Mash, 1989).

This is a 17-item scale within which there is a subscale of seven items that assesses efficacy. They are rated on a six-point response scale, ranging from 'Strongly agree' to 'Strongly disagree'. The items are positively worded, and some refer to a 'target' child although the wording has been adapted to refer to 'children' for some studies (Gilmore and Cuskelly, 2008). However, the wording of some of the individual items within the scale tended to be lengthier and more convoluted in comparison to other questionnaires, for example, 'The problems of taking care of a child are easy to solve once you know how your actions affect your child, an understanding I have acquired'; and 'I would make a fine model for a new mother/father to follow in order to learn what she/he would need to know in order to be a good parent'.

Parental Competence subscale of the Parenting Stress Index (PSI; Abidin, 1983). This scale has been used in previous research with Home-Start, and has been successful in demonstrating the impact of the service. It is a thirteen-item scale that uses a five-point response format, and has good reliability and validity (Abidin, 1995). It is one of the most widely used instruments for assessing parental competence in the field of family research (Jones and Prinz, 2005). While it has been shown to be a useful tool in a more extensive evaluation of Home-Start, it was not considered appropriate on this occasion for a number of reasons. It contains several items that refer to a 'target' child rather than all children within the family, and it has several negatively phrased items that focus on parenting problems rather than positive parenting experiences. It also asks mothers and fathers about their education level, and Home-Start Co-ordinators considered that this item might alienate parents who could potentially feel judged by the service's assessment. Another drawback for services wishing to use this measure as part of their evaluation is that there are costs associated with obtaining the measure and strict copyright restrictions that limit the format that it can be used in.

Appendix 3. Information sheet for parents inviting them to take part in the project



Information sheet for families

Please can you help us find out a simple way to ask you what family life is like for you?

Background

Home-Start Northern Ireland is involved in a project with the Centre for Effective Services that aims to work out a simple way of capturing what family life is like for you. We would like to ask you a few simple questions as part of our normal first and review visits over the next 9 months. It should only take between 5 and 10 minutes extra at each visit.

What about confidentiality?

Everything that you tell us will be treated with strictest confidentiality. The forms we complete will be stored in a locked cabinet when not being used and any information, written on paper or entered on a computer, will be anonymous (i.e. your name will be removed).

What will happen to the information that I give?

We would like to publish a report about the results, and no-one taking part will be identified or named in the report.

What if I change my mind about giving extra information for the research?

You can change your mind or decide not to take part at any stage. This will not affect the support you receive from Home-Start.

Will being involved affect the support I receive?

No. Whether you give additional information or not you will still receive the highest level of support from Home-Start.

What if I have any questions?

If you have any questions please contact any of these people:

your local Home-Start organiser/co-ordinator

Therese McCann Home-Start consultant

Appendix 4. Data collection process questions for Home-Start Co-ordinators

1. How easy was it to introduce and explain the project to parents?
Where there any particular difficulties you encountered in doing this?
2. How easy was it for parents to understand the purpose of the project?
Where there any particular difficulties they encountered in understanding the purpose?
3. How well do you think parents responded to being asked to fill in a questionnaire?
Where there any particular difficulties? (E.g. fears about what the data would be used for, literacy problems, time taken to fill it in, etc)
4. Did parents' attitude towards the request to fill in the questionnaire change on the follow-up visits? If there were any differences in attitude, what were they?
5. Did parents have any difficulty understanding the wording of the questions? If so, did that this happened often? Were any particular questions problematic?
6. Did you have any sense of parents trying to 'please' you (or 'fake good') or else reporting that things were substantially better than you thought they really were? If so, did this happen often?
7. Did you have any sense of parents trying to 'downplay' their coping in the way that they replied to the questions? (I.e. any parents that seemed to be trying to rate their coping as worse than it really was.) If so, did that this happened often?
8. Did you have any sense of parents just 'going through the motions' and ticking off any box without really bothering to take in what they were being asked? If so, did this happen often?
9. How easy or hard was it for parents to answer questions 4 and 5 in the parent questionnaire in relation to all of their children rather than just one child?
(A reminder of the items: 4. I can solve most problems between my child and me. 5. When things are going badly between my child and me, I keep trying until things begin to change.)
10. If you had to start this project all over again, what might you do differently?
11. Any other observations or thoughts about the data collection process?



Home-Start UK

Registered Office: the Home-Start Centre, 8-10 West Walk, Leicester, LE1 7NA

Tel: 0116 258 7920

Email: info@home-start.org.uk

Registered charity number: 1108837 (England and Wales), SC039172 (Scotland)

Registered Company: 5382181

www.home-start.org.uk

The Colebrooke Centre for Evidence and Implementation

55 St John Street

London EC1M 4AN

The Colebrooke Centre for Evidence and Implementation is a non-profit company limited by guarantee and registered in England. Company Number 07712883

www.cevi.org.uk

The Centre for Effective Services

9 Harcourt Street, Dublin 2, Ireland

and

Forestview, Purdy's Lane, Belfast BT8 7ZX, Northern Ireland

The Centre for Effective Services (CES) is a not-for-profit company limited by guarantee (Company Number 451580 and Charity Number 19438 in Ireland).

www.effectiveservices.org

© 2013 Home-Start UK, Deborah Gbate and the Centre for Effective Services

All rights reserved.

No part of this report may be reproduced without prior permission from the copyright holders.