



SOCIAL RESEARCH NUMBER:

70/2016

PUBLICATION DATE:

02/11/2016

Evaluability Assessment of Flying Start

Mae'r ddogfen yma hefyd ar gael yn Gymraeg.
This document is also available in Welsh.

Title: Evaluability Assessment of Flying Start

Author(s) / Company

Joseph Wilton, Welsh Government
Hannah Thomas, Welsh Government



Llywodraeth Cymru
Welsh Government

Views expressed in this report are those of the researcher and not necessarily those of the Welsh Government

For further information please contact:
Welsh Government
Cathays Park
Cardiff
CF10 3NQ
Tel: 0300 025 2663
Email: joseph.wilton@wales.gsi.gov.uk

Table of contents

1.	Summary	2
2.	Introduction	4
3.	The Flying Start Programme	5
4.	Programme Rationale	8
5.	Evaluability of Flying Start	15
6.	Currently Possible Approaches	25
7.	Potential Future Approaches	31
8.	Conclusions.....	34

1. Summary

- 1.1. This report explores the Flying Start programme and its rationale, how it can be evaluated and what might need to change for more robust and reliable evaluations to be conducted. Part of this is the 'theory of change' that underpins the programme, which provides details about its context, aims, activities and intended outputs and the expected outcomes of these outputs if achieved. This assessment also includes an evaluability checklist, which systematically explores all aspects of the programme to determine the degree to which it can be evaluated and in what respects. The completed checklist, combined with the findings from evaluations conducted to date, provides a clear understanding of what can currently be concluded with reasonable certainty about the effects of the Flying Start programme, and what cannot. At present, what can be, or already is, known about the programme is based on:
 - The differences between Flying Start areas and non Flying Start areas for some of the intended outcomes
 - The changes in some of the intended outcomes for Flying Start and non Flying Start areas after the intervention had been operating for a number of years
 - Parents' and stakeholders' perceptions of the programme, its specific entitlements and their perceived barriers to engagement
- 1.2. While these topics are of interest, and can provide useful information, they do not provide sufficient information to fully assess the effects of the programme on families, parents and children. There can also be no assessment of the indirect effects of the programme on wider services or family. Ideally, information for the following would be available:
 - Which children and families are receiving Flying Start entitlements
 - What entitlements these children are receiving, and how much
 - The characteristics of the children and families that engage with Flying Start
 - Individual level outcomes
 - Between those that make use of Flying Start and each particular entitlements and those that do not
 - Between those that have high levels of engagement and those with low levels
- 1.3. The main requirement for a more robust evaluation of the programme is to have data on families' engagement with Flying Start at the individual level. If this was collected then it would be possible to explore the effect of each entitlement, whether outcomes are dependent on the particular combinations of entitlements and/or individual or household characteristics and if there is a minimum or optimum level of engagement required for improved outcomes. Ideally this data would then be matched against data from other sources, such as the National Pupil Database, to track children who have received support from Flying Start when they start school.
- 1.4. This report makes the following recommendations for any future evaluations of the programme in its current form:
 - Compare outcomes between Flying Start and non-Flying Start areas using statistical matching to identify those in each area who have similar characteristics
 - Continue using qualitative research to understand the perceptions of families living in Flying Start areas
 - Explore options for data linking, using administrative data to potentially identify individual level outcomes for children living in Flying Start areas

- Identify a sample of families/parents/children who have engaged with Flying Start entitlements and track them longitudinally

The possible improvements to the programme's monitoring that could lead to more robust evaluations:

- Develop new systems for collecting individual level data about engagement with Flying Start entitlements, which could include either additional monitoring and/or a large scale survey
- Link individual level data collection to data collected by other services, such as schools' educational data
- Identify where families/parents/children are receiving additional services or interventions, and link this to the individual level data from Flying Start.

If these improvements were to be made, the following activities may now be possible:

- Using statistical matching, but at an individual level to determine the effects of specific Flying Start entitlements.
- Analyse the individual level data to see whether individual characteristics, or wider contextual factors, affect take-up and impact of entitlements.
- Use individual level data for selecting samples for further qualitative research and/or more in-depth surveys, case studies and longitudinal research.
- Identify where additional services or interventions (not Flying Start) are having an additional effect on those in Flying Start areas

2. Introduction

- 2.1. The aim of this report is to set out the issues and challenges that arise when attempting to conduct a robust evaluation of the Flying Start programme. Several evaluations of the programme have been conducted since 2009 with a variety of methodologies. The first was commissioned in 2007, but crucially, this evaluation was developed after the implementation of the programme. Each subsequent evaluation has referred to some of the problems of evaluating Flying Start, but to date there has been no coherent evaluability assessment of the programme.
- 2.2. The term evaluability is described as *“The extent to which an activity or project can be evaluated in a reliable and credible fashion”* (OECD-DAC 2010; p.21). Any assessment of the evaluability of a programme will attempt to determine how successful any evaluations will be in providing useful and reliable evidence.
- 2.3. How this assessment is conducted depends on its timing relative to the programme or policy being assessed. As this assessment is being conducted several years after implementation, the purpose is to:
 - a) inform future evaluations and research,
 - b) potentially shape the future of the ongoing data collection and evaluation strategy to allow for more effective evaluations and to,
 - c) determine if future evaluations are feasible, and what approaches they could take.
- 2.4. The Department for International Development (DFID) produced a checklist which can be used to ensure all aspects of a programme are considered in a systematic manner. The output of this checklist can be used to determine how effective any evaluation is likely to be, and to identify the areas in which the programme can be changed or improved. This checklist is used in section 4.
- 2.5. In general, the main questions which should be asked of any programme¹ are:
 - Is it plausible to expect impacts?
 - Is it feasible to measure impacts?
 - Would an impact assessment be useful?
- 2.6. In order to answer these questions, it is necessary to have a full understanding of the programme. This includes how the programme is being delivered, the context in which it operates, the aims and objectives of the programme and the rationale for the programme.
- 2.7. Section 3 sets out the Flying Start programme and what the guidance states should be offered to all eligible children under four years of age living in Flying Start areas. Section 4 describes the programme in more detail and explores the rationale and logic behind Flying Start entitlements. Section 5 applies the DFID checklist to determine what aspects of the programme allow for an effective evaluation, and what aspects would need to be changed. Section 6 describes the methods and evaluation approaches that are currently possible. Section 7 then explores what approaches could be possible if changes were made, and what advantages and costs are associated with these changes.

¹ http://pdf.usaid.gov/pdf_docs/PNADN200.pdf

3. The Flying Start Programme

3.1. Flying Start is the Welsh Government's flagship Early Years programme for families with children who are under 4 years of age. It is aimed at improving the life chances of children living in some of Wales's most disadvantaged communities and is one of the top priorities in the Welsh Government's Tackling Poverty Agenda. The 4 key entitlements of the Programme are:

- Free quality, part-time childcare for 2-3 year olds
 - Flying Start provides quality childcare which is offered to parents of all eligible 2-3 year olds for 2 and a half hours a day, 5 days a week for 39 weeks a year. In addition, there should be at least 15 sessions of childcare for the family during the school holidays.
 - This should be linked to Foundation Phase entry into schools to ensure a seamless transition between the two offers with no gaps in provision.
- An enhanced Health Visiting service
 - Key to Programme delivery is the requirement that there must be one full time equivalent Health Visitor per 110 children aged under 4 in the target areas. This is to ensure delivery of intensive support to Flying Start children and their families.
 - The primary function of the Flying Start Health Visitor is to support the family in the home, assessing both the child and the family (in terms of risk to the child's health and development). Flying Start Health Visitors should continually assess those families identified as medium and high risk, and make appropriate referrals.
- Parenting support
 - Every family with a Flying Start child must be offered both formal and informal parenting support at least once a year². This can be in groups or one to one in the home with a mix of formal and informal support depending on need.
 - The parenting offer should be based on provision of perinatal and support in the early years to age 4. This should be underpinned by the following cross-cutting themes:
 - Relationship support
 - Early intervention to support vulnerable families
 - Positive parenting
 - Evidence based theories of child development
- Speech, Language and Communication support
 - Every family in a Flying Start area should have ongoing access to an appropriate language and play group. From this, a more targeted approach based on assessment and referral can be taken where there is evidence of additional need. In some local authorities, speech and language therapists are employed as part of the core Flying Start team.
 - Evidence based key messages to support children's speech, language and communication have been developed which should be shared with all parents by Flying Start teams and embedded throughout all entitlements of the programme.

3.2. The Flying Start programme was launched in 2006/07 and has expanded in its coverage over time. In 2012/13, some 23,500 children received support from Flying Start. There is a continuing commitment from the Welsh Government to increase the number of children receiving Flying Start to 36,000 by 2016³. During 2013-14 an expansion of the

² This support includes the following: Evidence-based, group-based structured parenting programmes, One-to-One support, Informal Structured Group-based Parenting Support and Informal Drop-in Support

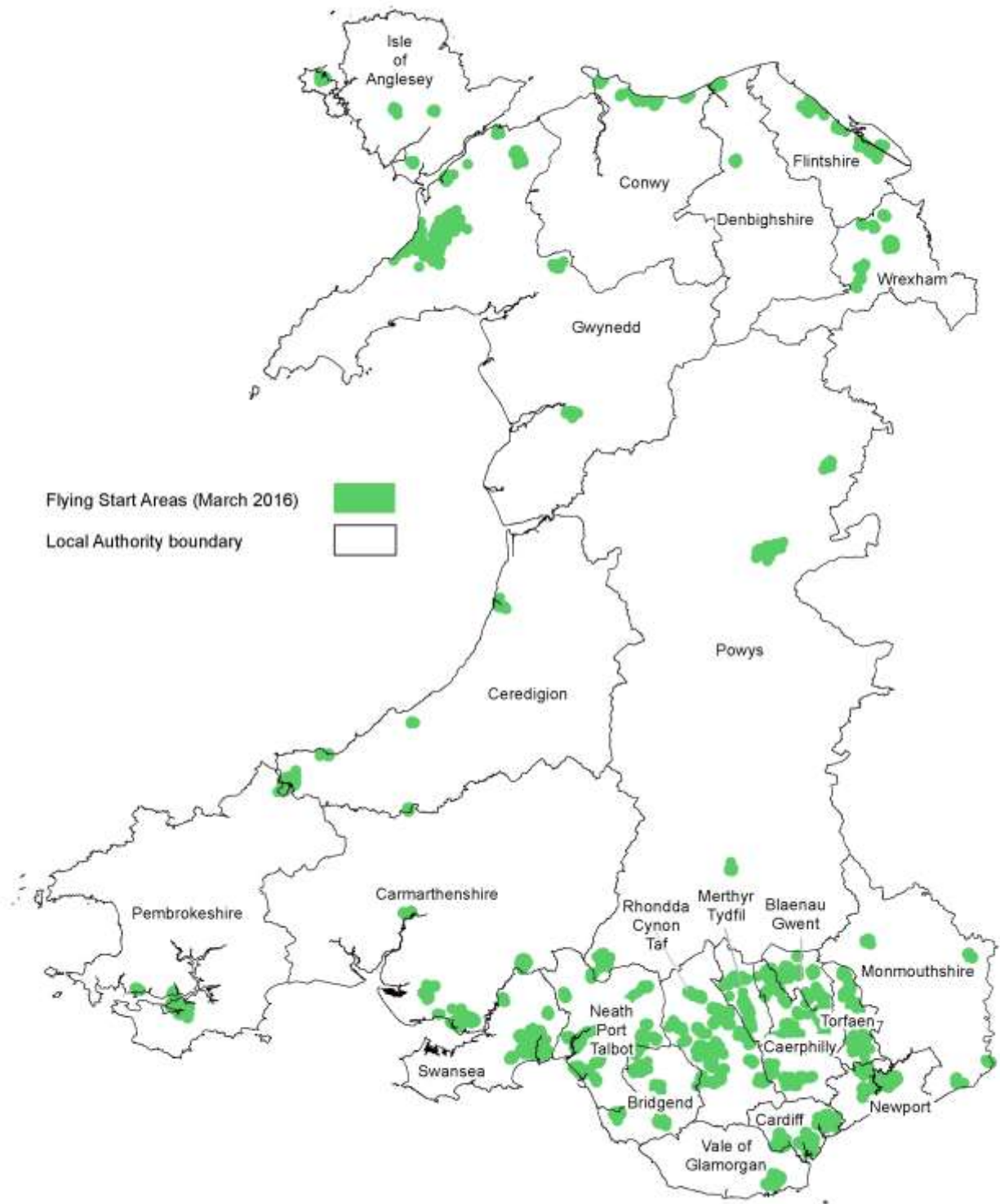
³ <http://www.assemblywales.org/RN14-005.pdf>

programme into additional areas began, continuing in 2014-15. The latest statistics indicate this commitment has been achieved, as Flying Start is now being received by just over 38,000 children, which exceeds the expected number.

- 3.3. In addition to the 4 core elements, Local Authorities are also able to apply a degree of flexibility within the Flying Start programme by offering support through outreach. Outreach enables a small number of families living outside Flying Start areas to access the support they need. Using local knowledge and an assessment of priority, Local Authorities can aim to ensure those most in need receive this service.
- 3.4. The evaluation of Flying Start is ongoing and has taken a mixed methods approach, encompassing a large-scale survey, longitudinal case studies with Flying Start families and in depth qualitative research with 'high need' families⁴. The evaluation of the Flying Start programme has also formed the basis of a 'Data Linking Demonstration Project'. The study took Flying Start as an exemplar of the issues and potential benefits associated with the use of administrative data for the purposes of research. Focussing predominantly on health based measures, the project demonstrated that the introduction of Flying Start appeared to be associated with improvements in a variety of outcome measures⁵.
- 3.5. Evaluation of the Flying Start programme to date has included surveys with families in both Flying Start areas and areas that were most similar in terms of area deprivation levels, but which themselves were not Flying Start areas. The aim of this was to determine the impact of the programme, by using those in the similar areas as a counterfactual. The problems with such an approach is that Flying Start was first introduced in the most disadvantaged areas of Wales, which inevitably means that comparisons made with children outside of the Flying Start areas will result in comparisons being made with children in relatively less disadvantaged areas. The map below shows the areas, as of 2016, which were in receipt of Flying Start entitlements.

⁴ See <http://wales.gov.uk/statistics-and-research/national-evaluation-flying-start/?lang=en> for an overview of the evaluations of Flying Start.

⁵ <http://wales.gov.uk/statistics-and-research/data-linking-demonstration-projects/?lang=en>



Flying Start Areas (March 2016)

Local Authority boundary



4. Programme Rationale

- 4.1. In evaluating a policy or programme, it is critical to develop a full understanding of the theory behind it. This means clearly defining the steps involved in delivering the policy, the context in which it is operating, what resources are required, what is actually undertaken as part of the policy and what changes are expected as a result. One way in which this understanding can be developed and described is through a logic model.
- 4.2. The Magenta Book states '*Logic models describe the relationship between an intervention's inputs, activities, outputs, outcomes, and impacts*' (pg. 22).
- Inputs are the quantifiable resources that are required to deliver the policy/programme. This can include funding, staff, buildings or physical materials, such as books or leaflets.
 - Activities are the actual tasks involved in delivering the policy. For example, this could be providing training courses, distributing information or building a road.
 - Outputs are the direct results of the activities and what the recipient does as a response to the activities. So in the example of providing a training course, the output could be people attending the course.
 - Outcomes are results of the outputs and will occur some time after the activity. Continuing with the example of a training course, the outcome would ideally be that the recipients' skills or knowledge are increased.
 - Impacts are the long-term, wider effects of a policy/programme, and are usually considered the ideal goal. They are often considered on a population level and so would describe how and activity can lead to lasting changes in the group involved in the activity. On a national level this may be incomes are increased, or health improves.
- 4.3. A logic model can also be thought of as a means of portraying a theory of change. A theory of change attempts to explore the assumptions around a programme and on what basis it claims to be able to bring about the outcomes and impacts that are stated in the logic model⁶. It can be broadly thought of as a critical appraisal of a logic model, by which the links between the inputs, activities, outputs, outcomes and impacts are explored and the assumptions are stated and challenged.
- 4.4. Inherent in all social policies are assumptions about how the policy will engage with the target recipients. These can range from the size of the expected recipient population to how the activities will influence the recipients outcomes. It is these assumptions that are tested in policy evaluations.
- 4.5. In addition to the underlying assumptions, a theory of change also sets out the context in which the policy or programme is operating. This is a vitally important part of any evaluation, as the context can have a big influence on the effect of an intervention. The context in this case refers to the political, social and economic systems that are in place, the situation of the target beneficiaries prior to implementation (i.e. the baseline) and who is involved in the both the implementation of the programme and the problem being addressed. It should also include the mode of implementation, which gives an indication of the programme's fidelity.

⁶ http://r4d.dfid.gov.uk/pdf/outputs/mis_spc/DFID_ToC_Review_VogelV7.pdf

- 4.6. The 2013 Impact Report⁷ presented a model for Flying Start which articulated the rationale for the programme, the context, change assumptions and the various elements required for a logic model. While this was a useful illustration of the key points of the programme, it fails to capture all the various elements in detail. The Flying Start programme is highly complex, with many interacting services and entitlements, which were not fully illustrated in the 2013 model. Therefore an expanded and updated model has been produced.
- 4.7. This new model attempts to account for the context in which the programme was developed and implemented, the key aims and objectives of the programme and how it is being delivered. This model attempts to provide a logic model with the roles of the four key entitlements clearly articulated. There is also a consideration of the main issues and challenges that can be faced at each level of the model. These issues are an attempt to explore the assumptions, and the potential for these assumptions to hold true or be poorly founded.
- 4.8. While every attempt has been made to make this model as detailed and exhaustive as possible, it needs to be acknowledged that the programme is delivered in different ways in each Local Authority⁸. Therefore this model represents the national delivery of Flying Start, and may not be applicable to each local area.
- 4.9. In addition to local variations, the programme is subject to change in delivery over time. For example, the speech, language and communication entitlement underwent a large scale redesign as a result of the 2013 evaluation. Therefore, this model may cease to be relevant or accurate if further changes are made.

⁷ <http://gov.wales/docs/caecd/research/131205-national-evaluation-flying-start-impact-report-en.pdf>

⁸ Whilst the programme is prescriptive about what entitlements to provide, the differences in local availability of services, staffing and recipient characteristics result in variation in each Local Authority.

Context

Political priorities at the time of programme development focussed on prevention and early intervention, helping children develop and to support families. Political priorities shifted over the course of the programme's life, with topics such as education, health and poverty gaining and losing focus. This has consequential effects on how the programme was delivered. However, the focus and delivery of the programme has not changed since the start of expansion in 2012.

Tackling child poverty has been a key goal of the Welsh Government for many years, and the first Child Poverty Strategy for Wales was published in 2005. One of the aims of this strategy was to eradicate child poverty by 2020. In 2005 nearly a third of children in Wales were considered to be living in households in relative poverty. The evaluation of the Child Poverty Strategy for Wales describes how the wide reaching effects of poverty on children's development have been recognised for some time, and so there was a push to prevent and investigate child poverty. Early years interventions were viewed as potentially very good value for money because the benefits are accrued for a long time after the intervention, and can lead to prevention of future gaps in attainment and negative outcomes.

There are many factors which can influence a child's development, from the most immediate (i.e. family), to the local area (e.g. neighbourhood) to wider structural forces (e.g. school system, welfare system). A policy which aims to tackle all these factors is potentially more effective than one which targets single factors. Many can be directly influenced by Welsh Government, some need support from UK government actions (e.g. benefits, taxes) whilst others are sub-national (e.g. Local Authorities), see [Child Poverty Strategy for Wales](#).

Aims and Objectives of Flying Start

The overall aim of Flying Start is to reduce the impact of deprivation on children by investing in early year's development
'to make a decisive difference to the life chances of children aged under 4 in the areas which it runs'
<http://gov.wales/docs/dhss/publications/120913fsguidanceen.pdf>, pg. 3

While the Flying Start Strategic Guidance is very clear on how the programme should be delivered, it does not include any specific or testable objectives. However, the various evaluations to date have highlighted several possible objectives of the programme. These objectives have been distinguished between those that are for the children themselves, the parents and for the delivery of the service.

Children Objectives

- Language development
- Cognitive development
- Social and emotional development
- Physical health
- Early identification of high needs

Parents Objectives

- Support for parents
- Positive parenting behaviours
- Parenting confidence

Wider Objectives

- Expanded workforce
- Take up of core entitlements
- Use of wider services
- Professional collaboration
- Infrastructure

Delivery of Flying Start

Flying Start was developed to provide targeted support for families with children under the age of 4 living in the most disadvantaged areas in Wales. This targeting was based on the scores produced through the Welsh Index of Multiple Deprivation but also on data from HMRC and DWP.

The programme was officially implemented from 2006/07 and was then expanded in 2012/13. The aim of this expansion was to double the number of children eligible for the programme from the initial 18,000 to 36,000 by 2016.

An outreach element was also introduced in 2012/13, which takes up about 2.5 per cent of the increase in each Local Authority's annual allocation of Flying Start funding on a cumulative basis.

Inputs

£76m provided annually to Local Authorities – distributed by population – roughly £2,100 per child

This does not include capital costs, such as those used to build new settings and the SoGS assessment forms and training.

Local Authorities are given freedom to distribute their funds between the four entitlements

Health

- Guidance for core health programme
- Health visitor ratios
 - One per 110 children
- SoGS assessment
 - Forms, training
- Access to specialist services
 - Specialist health workers (e.g. psychologists)
 - Greater local area flexibility in deciding work force skills mix

Childcare

- Higher quality standards imposed than minimum standard
 - Lower ratios
 - Higher qualifications including at least one degree-level staff member
 - Advisory teacher
- Capital funding
- Full CSSIW registration required

Speech, language & communication (SLC)

- SLC key messages
 - Expectation all FS staff use these messages pre-natal onwards (e.g. childcare staff, health visitors)
- Evidence based guidance
- Flying Start Book Bags
- Posters for childcare settings
- "Learning To Talk" Parent Packs
- Speech and Language Development Wheels for health visitors

Parenting

- Guidance issued on approved evidence based parenting courses and informal courses
- Flying Start health visitors and parenting practitioners

Issues and Challenges: Health visitor ratios not achieved in all areas, specialist services vary by area and can depend on local availability, need and supply of staff, sourcing sufficiently qualified childcare staff not always achieved, SLC key messages introduced in 2015, SLC resources changed over lifespan of Flying Start, parenting programmes at discretion of local area within guidance expectations

Activities

Core programme of intervention – should be provided as a minimum
Local Authorities able to provide additional activities based on area and family need
Some collaboration with Families First and Communities First

Health

- SoGS assessments
 - 2 and 3 year old children assessed
- Health visitor contacts
 - 13 core contacts plus additional when required
 - Developmental tracking
- Health visitor training
 - Specific to Flying Start (e.g. in SoGS)
 - Identify needs and risks
- Health related courses provided
 - E.g. healthy cooking, breast feeding
- Healthy choices promoted
 - E.g. immunisations, dentist visits

Childcare

- 2.5 hours/weekday (12.5 hours/week) for 39 weeks
 - Offered to 2 to 3 year olds
 - Additional to Foundation Phase Early Years Education
- 15 sessions available during school holidays
- Staff recruited at higher qualifications than minimum standard
- Additional training (e.g. in SoGS, SLC)
- New settings built
- School transition support
 - Handover of developmental journal
 - Introduce child to new setting

Speech, language & communication (SLC)

- SLC resources distributed (e.g. Flying Start Book Bags) to parents and staff
- SLC key messages embedded throughout entitlements
- Training on key messages
 - Each area has own approach
 - Childcare staff, health visitors etc.
- Language and play groups
 - Encouraged by health visitors
- Health visitors give SLC key messages to families
 - Improve parents' ability/confidence
 - Monitor child development
- Referrals to Speech and Language Therapists where necessary

Parenting

Begins antenatally:

- Parenting courses
 - Areas provide approved evidence based courses
- Informal parenting support
 - Includes health visitor contacts, group/one-to-one sessions
 - Varied according to local area
 - Aimed at meeting specific family needs
 - Practical and emotional support
 - Aimed at being convenient for parents
- Parenting groups
 - Peer-to-peer support

Issues and Challenges: Health visitors can provide as many contacts as they deem required for the family, health visitor contacts constrained by case load, available support may not be address family's true needs (e.g. employment skills, poverty), childcare take-up is not 100%, staff recruitment depends on local availability of staff, consistency of SLC activities across areas has been identified as an issue for Flying Start in previous evaluations, greater emphasis on SLC introduced in 2015.

Outputs

Local areas are required to submit data for monitoring purposes to the Welsh Government which gives area level information for many of these outputs

Outputs vary by local area

Individual level data is not available for these outputs

Many of these outputs require engagement from staff across entitlements, additional services and/or parents

Health

- Children assessed
- Children with developmental delays identified
 - Early identification and intervention
 - Referrals to specialists
- Health visitors provide specialised support
- Children's development tracked
- Family risk factors identified
 - Support given
- Children receive immunisations
- Breast feeding courses attended
- Parents receive advice

Childcare

- Children attend childcare
 - Absenteeism chased up with families
- Quality staff recruited
- New settings used
- Settings inspected
- Transitional work undertaken
 - Developmental issues passed on to new setting
- Children receive developmental support
 - Targeted at children with identified issues

Speech, language & communication (SLC)

- SLC activities undertaken and encouraged
- Resources received
- Staff attend SLC training
- Reading groups attended
- Language and play courses attended
- SLC activities used in childcare settings

Parenting

- Parenting courses attended
- Parenting groups attended
- High need families receive focused support
- Parenting needs identified

Issues and Challenges: SoGS assessments vary across local areas and by assessors (can be conducted by any trained individual including childcare staff), early intervention depends on available specialist services, levels of attendance for courses unknown, family diet change unknown, dentist visits unknown, childcare take up optional, new settings only in some Flying Start areas, use of SLC resources unknown, reading groups depend on local area support, unclear if all high need parents receive the necessary support

Outcomes (0 to 1 year after activity)

Health

- Children with developmental delays, and their parents, receive specialist support to either reduce impact of, or prevent, long term conditions
- Home life conditions and children's general health improve
- Families' diets improved
- Fewer children with social or emotional behavioural difficulties

Childcare

- Children become accustomed to attending a formal care setting
- Children interact with other children of own age
- Children's cognitive and social development improves
- Families become engaged with learning and development
- Inequality in learning opportunities between areas reduces

Speech, language & communication (SLC)

- Parents understand importance of SLC key messages
- Parents engage with children's learning and use SLC activities
- Children's SLC skills develop and improve
- Children develop ability to learn through play and so enjoy process of learning
- Home learning environment improved

Parenting

- Antenatal support taken up by parents
- Number of risky behaviours reduced in households (e.g. smoking)
- Parents have greater understanding of their role in their children's development and learning
- Children's home life is more supportive and stimulating

Issues and Challenges: Some development delays identified by SoGS will specialist intervention from wider health service and parental engagement, some conditions cannot be identified before the age of 4, specialist support may not be available, some aspects of children's health not treatable through education and encouragement (e.g. fuel poverty), childcare may not have any beneficial effect on development, parents may not engage with children's development, Flying Start provision may not be sufficient to combat inequalities, SLC development depends on parental engagement, parenting groups may not be sufficient to change behaviour, highest risk groups may not engage

Impact (1 to 5 years after activity)

Child

- Educational outcomes for primary school children in disadvantaged areas improved
- More children meeting expected levels in literacy throughout school
- Improved social behaviours
- Greater developmental opportunities
- Fewer children being seen by medical services (GPs, hospital etc.)

Parents/Family

- Fewer families requiring social service interventions
- Reduced number of parents needing additional support
- Fewer parents being referred for physical or mental health problems
- More parents feeling confident in their ability to raise children

Service

- Higher numbers of qualified childcare workers
- Greater engagement between early years support, schools and other social interventions (e.g. Families First)
- Reduced costs for additional support in schools
- Lasting infrastructure improvements

5. Evaluability of Flying Start

- 5.1. One of the key aspects of the evaluability assessment is the checklist developed by the DFID. This checklist has been extracted from pages 19-23 of the following report: Davies, R., 2013. *Planning Evaluability Assessments: A Synthesis of the Literature with Recommendations*. Report of a study commissioned by the DFID.
- 5.2. The evaluability assessment, in the form of the checklist, identifies the following dimensions of evaluability;
- Evaluability “in principle”, given the nature of the project theory of change
 - Evaluability “in practice”, given the availability of relevant data and the capacity of management systems able to provide it.
 - The utility and practicality of an evaluation, given the views and availability of relevant stakeholders
- 5.3. The checklist aims to provide an accountable means of ensuring coverage of all the relevant issues through a systematic process. The expected assessment outputs of this checklist concern not only the evaluability of the project but also the practicality and utility of an evaluation. In addition, a secondary set of outputs will be recommendations in terms of how to make the project more evaluable.
- 5.4. Table 1 shows the completed evaluability checklist for the Flying Start programme. In completing this evaluability assessment, it is hoped this will inform future evaluations and research and potentially make suggests for how the programme could change to allow for more effective evaluations.

1. Project Design (as described in a Theory of Change, Logical Framework or narrative)		Notes – (positives and <i>negatives</i>)
Clarity?	Are the long-term impact and outcomes clearly identified and are the proposed steps towards achieving these clearly defined?	<ul style="list-style-type: none"> • Primary aims and key focus has been the same: the development and wellbeing of children in disadvantaged areas • While there has been some variation in how the activities operate, the main components and approach is the same
		<ul style="list-style-type: none"> • <i>The secondary aims of the intervention have changed focus and priority, such as trying to show impact on parental employment</i> • <i>While some indicators are reported at an aggregate level, the intended effects of the programme are not very clearly defined, i.e. in terms of specific outcomes or targets.</i>
Relevant?	Is the project objective clearly relevant to the needs of the target group, as	<ul style="list-style-type: none"> • Beneficiary group identifiable through living in Flying Start area. Areas are clearly defined and selected using reliable data

	<p>identified by any form of situation analysis, baseline study, or other evidence and argument? Is the intended beneficiary group clearly identified?</p>	<ul style="list-style-type: none"> • Standard format of provision • Intervention is bespoke to each family <hr/> <ul style="list-style-type: none"> • <i>Unable to determine at the individual level the entitlements used by beneficiary group</i> • <i>The target group are defined as 'at risk', rather than definitely being in need of the intervention</i> • <i>Flying Start includes a small outreach component, which provides Flying Start entitlements to those outside of the Flying Start areas</i> • <i>There is no clear comparison between the services available in Flying Start and non Flying Start areas. Some of the Flying Start entitlements may also be available in non Flying Start areas, although through other providers.</i>
<p>Plausible?</p>	<p>Is there a continuous causal chain, connecting the intervening agency with the final impact of concern?</p> <p>Is it likely that the project objective could be achieved, given the planned interventions, within the project lifespan? Is there evidence from elsewhere that it could be achieved?</p>	<ul style="list-style-type: none"> • Intervention based on robust evidence <ul style="list-style-type: none"> ○ Programme logic supported by evaluations of previous initiatives, such as Sure Start⁹ ○ Childcare entitlements based on strong longitudinal evidence, such as that produced through the Effective Provision of Pre-School Education study¹⁰ ○ Formal parenting courses only recommended if they have robust supporting evidence ○ Speech, language and communication activities based on sound evidence as recommended by qualified Speech and Language Therapists. • Evidence used at programme development stage suggests providing more support, improved access or services at no cost to the families receiving the services is likely to lead to positive impacts <hr/> <ul style="list-style-type: none"> • <i>Project lifespan extends from birth to when the child is 4 years old. Some impacts are likely to be seen well beyond this range, such as into child's school years and potentially adulthood</i> • <i>While evidence supports individual activities and components of Flying Start, the effect of the interaction of entitlements is unknown</i> • <i>Some of the underlying evidence for the programme was produced in other countries (mainly the USA). Therefore may not be generalisable to a Welsh context.</i>

⁹ https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/182026/DFE-RR067.pdf

¹⁰ <http://eppe.ioe.ac.uk/eppe/eppepdfs/bera1.pdf>

Validity and reliability?	Are there <i>valid</i> indicators for each expected event (output, outcome and impact levels)? I.e. will they capture what is expected to happen? Are they <i>reliable</i> indicators? I.e. will observations by different observers finding the same thing?	<ul style="list-style-type: none"> • Some data exists from currently collected administrative sources, such as educational assessment scores, school absenteeism, immunisation and breastfeeding rates • Able to compare between Flying Start and non Flying Start areas, and against Flying Start areas prior to implementation of the programme
		<ul style="list-style-type: none"> • <i>Unable to measure output (i.e. take up of entitlements) at the individual level, and so cannot determine whether any change could be explicitly attributable to Flying Start entitlements.</i> • <i>The objectives of Flying Start were not specified at implementation, and so potential outcomes and impact were not identified.</i> • <i>There is some question over the coverage and reliability of the administrative data sources</i>
Testable?	Is it possible to identify which linkages in the causal chain will be most critical to the success of the project, and thus should be the focus of evaluation questions?	<ul style="list-style-type: none"> • All families living in Flying Start areas should have access to core entitlements, such as reduced case-load health visitors, parenting support, SLC and childcare
		<ul style="list-style-type: none"> • <i>Flying Start adopts a holistic, whole family, approach and so some interventions delivered are dependent on family need. This means the provision is different for each family that receives entitlements</i> • <i>Due to the lack of individual level data, there is no ability to test the linkage between the entitlements used by those in Flying Start areas</i>
Contextualised?	Have assumptions about the roles of other actors outside the project been made explicit? (both enablers and constrainers) Are there plausible plans to monitor these in any practicable way?	<ul style="list-style-type: none"> • Local authority responsible for allocation of funds to Flying Start entitlements, and their role in delivering the programme is clearly understood
		<ul style="list-style-type: none"> • <i>The way in which Local Authorities deliver Flying Start varies according to local factors, levels of demand, available specialist services etc. and continues to change over time. There is currently no comprehensive data collection on the specific differences between local delivery</i> • <i>Several contextual changes have occurred during the lifespan of the programme:</i> <ul style="list-style-type: none"> ○ <i>Introduction of Healthy Child Wales</i> ○ <i>Introduction and changes to Families First and Communities First</i> ○ <i>Loss of Cymorth</i> ○ <i>Change in early years curriculum</i> ○ <i>Changes to benefits entitlements</i>

<p>Consistent?</p>	<p>Is there consistency in the way the Theory of Change is described across various project multiple documents (Design, M&E plans, work plans, progress reports, etc.)</p>	<ul style="list-style-type: none"> • Broad understanding of the programme is consistent, with the emphasis on children's development and well-being • Specific elements of the Flying Start programme have detailed Theories of Change and underpinning logic, such as the evidence based parenting courses
<p>Complexity?</p>	<p>Are there expected to be multiple interactions between different project components? [complicating attribution of causes and identification of effects] How clearly defined are the expected interactions?</p>	<ul style="list-style-type: none"> • <i>There was no global Theory of Change produced at inception, project development or implementation stages</i> • <i>Flying Start was originally conceived as a pilot, but the perceived early success led to it being rolled out and maintained</i> • <i>There was a period of 'bedding in' where the programme was not providing the full offer in all areas for roughly two years after implementation</i> • <i>Specific interventions have changed as a result of evaluation recommendations, such as the movement from Language and Play to Speech, Language and Communication support</i> <ul style="list-style-type: none"> • <i>Flying Start is a highly complex intervention with definite links between the different entitlements, which likely causes interactions between them e.g. it is possible the effectiveness of the childcare entitlement is dependant on the wider family needs being addressed by the health visitors</i> • <i>The effect of these interactions is unknown, and it is likely the bespoke, holistic nature of the programme results in a different take up of each entitlement and effect for each family, and consequently the interactions will be different and unknown</i> • <i>Families involvement in Flying Start is likely to increase the likelihood of them being signposted to other specialist services and being supported to claim additional benefits and services, such as housing support etc.</i>
<p>Agreement?</p>	<p>To what extent are different stakeholders holding different views about the project objectives and how they will be achieved? How visible are the views of stakeholders who might be expected to have different views?</p>	<ul style="list-style-type: none"> • Stakeholders and delivery bodies have a consistent view of the main aims of Flying Start • Local Authorities have clear targets regarding delivery. • There is a high level of buy-in from stakeholders, and the programme (and elements of the programme) were developed in collaboration with stakeholders <ul style="list-style-type: none"> • <i>There is Local Authority variation in the interpretation of the Flying Start guidance, availability of services and family needs, and therefore delivery will also vary</i>

2. Information availability		
Is a complete set of documents available?	...relative to what could have been expected? E.g. Project proposal, Progress Reports, Evaluations / impact assessments, Commissioned studies	<ul style="list-style-type: none"> Complete documentation available since the expansion of Flying Start, e.g. evaluation reports, monitoring returns
		<ul style="list-style-type: none"> <i>Early documentation about the initial project inception and development unavailable, possibly due to IT and record management changes</i> <ul style="list-style-type: none"> <i>Early documentation unlikely to be relevant to the current programme due to the changes between the original plans and current provision</i>
Do baseline measures exist?	<p>If baseline data is not yet available, are there specific plans for when baseline data would be collected and how feasible are these?</p> <p>If baseline data exists in the form of survey data, is the raw data available, or just selected currently relevant items? Is the sampling process clear? Are the survey instruments available?</p> <p>If baseline data is in the form of national or subnational statistics, how disaggregated is the data? Are time series data available, for pre-project years?</p>	<ul style="list-style-type: none"> Administrative data exists for children living in Flying Start areas prior to the implementation of the programme <ul style="list-style-type: none"> This data exists in both individual forms (e.g. educational attainment) and at an aggregate level (e.g. levels of deprivation)
		<ul style="list-style-type: none"> <i>Available administrative data is limited in topic and cannot provide a full understanding of the circumstances of the families living in Flying Start areas at implementation. It also only acts as a baseline of those eligible for Flying Start, not necessary those that received it.</i> <i>No survey of families was conducted at implementation, and it would be impossible to retrospectively survey these families due to them not being identifiable.</i> <i>Baseline of children entering the programme is inappropriate due to entry either being through birth or movement into a Flying Start area</i> <ul style="list-style-type: none"> <i>Those who move into a Flying Start area come from a range of circumstances, and so do not represent a consistent population</i>
Is there data on a control group?	Is it clear how the control group compares to the intervention group? Is the raw data available or just summary statistics? Are the members of the control group identifiable and potentially contactable? How frequently has data been collected on the status of the control group?	<ul style="list-style-type: none"> Best possible comparison groups are the areas that were included in the expansion of the programme. These areas are considered the next most disadvantaged and so are more similar to the original areas than those that receive no Flying Start services.
		<ul style="list-style-type: none"> <i>As Flying Start was implemented in most disadvantaged areas of Wales there is no equivalent (control) areas in Wales or elsewhere in the UK</i> <i>The nature of the programme does not lend itself to a randomised control trial, as specific area demographic conditions are required. The small number of these areas means they vary in quality to the point of lacking comparability</i>

Is data being collected for all the indicators?	Is it with sufficient frequency? Is there significant missing data? Are the measures being used reliable i.e. Is measurement error likely to be a problem?	<ul style="list-style-type: none"> Local Authorities provide the Welsh Government with some aggregate data on take-up of entitlements, such as number of developmental assessments completed, percentage of children attending childcare
		<ul style="list-style-type: none"> <i>This aggregate data is not available throughout the lifespan of the programme</i> <i>The indicators were not specified from the outset, so have been developed in response to the data that can be collected, rather than the preferred route of developing indicators and arranging data collection to measure against them.</i>
Is critical data available?	Are the intended and actual beneficiaries identifiable? Is there a record of who was involved in what project activities and when?	<ul style="list-style-type: none"> There is some data on individual children's developmental assessments, but it is not linkable to any other data source that includes the children who underwent the assessment.
		<ul style="list-style-type: none"> <i>There is currently no available national data on what entitlements have been used by individuals in Flying Start areas. Some Local Authorities collect accurate data on involvement, but this is inconsistent.</i> <i>The actual beneficiaries are not clearly defined. The aim is to improve the lives of children, but this may be via parents and other family members.</i> <i>Both children and families can be partial beneficiaries, who make use of some entitlements and/or for less time than the full eligible period.</i>
Is gender disaggregated data available?	In the baseline? For each of the indicators during project intervention? In the control group? In any mid-term or process review?	<ul style="list-style-type: none"> There is some aggregate data which specified between genders, such as health visits.
		<ul style="list-style-type: none"> <i>The lack of data means no disaggregation is available at the individual level</i>
If reviews or evaluations have been carried out...	Are the reports available? Are the authors contactable? Is the raw data available? Is the sampling process clear? Are the survey instruments available?	<ul style="list-style-type: none"> All previously conducted evaluations are available online on the Welsh Government website. Each has the details of the authors. Of those evaluations that included surveys, the sampling process is clear The survey instruments are available
		<ul style="list-style-type: none"> <i>The raw data for surveys prior to 2013 is not available</i>
Do existing M&E systems have the capacity to deliver?	Where data is not yet available, do existing staff and systems have the capacity to do so in the future? Are responsibilities, sources and periodicities defined and appropriate? Is	<ul style="list-style-type: none"> <i>Currently, only Local Authority level aggregate data is available on a limited number of fields. This data is not fit for purpose and needs to be reviewed.</i> <i>Local Authorities are aware of their responsibility for collecting monitoring data, but there is variation in the monitoring processes between Local Authorities due to their flexibility over resource</i>

	the budget adequate?	<p><i>allocation</i></p> <ul style="list-style-type: none"> • <i>Data is not available at the individual level and for all the entitlements in which the children and family can engage with.</i> • <i>There are concerns that there is insufficient resource (both staff and budget) to improve monitoring systems.</i>
3. Institutional context		
Practicality		
Accessibility to and availability of stakeholders?	Are there physical security risks? Will weather be a constraint?	<ul style="list-style-type: none"> • Little / no security risks in meeting with or contacting stakeholders, e.g. Local Authorities • All stakeholders are easily contactable / reachable
	Are staff and key stakeholders likely to be present, or absent on leave or secondment? Can reported availability be relied upon?	<ul style="list-style-type: none"> • <i>Locating and identifying recipients of Flying Start (e.g. parents) can be more difficult but is rarely required by Welsh Government directly. Recipients would have to be contacted via Local Authorities.</i>
Resources available to do the evaluation?	Time available in total and in country? Timing within the schedule of all other activities? Funding available for the relevant team and duration? People with the necessary skills available at this point?	<ul style="list-style-type: none"> • There is limited funding set aside for evaluation of Flying Start • Evaluations which have been done previously have provided recommendations and the programme has used the findings to inform policy development for future delivery
		<ul style="list-style-type: none"> • <i>Funding for future evaluations uncertain as the budgets are highly influenced by Ministerial priorities.</i> • <i>There is a drive for future evaluations to make use of existing data, rather than surveys or any bespoke data collection. However, detailed analysis of this data requires technical expertise which may not be available.</i>
Is the timing right?	Is there an opportunity for an evaluation to have an influence? Has the project accumulated enough implementation experience to enable useful lessons to be extracted? If the evaluation was planned in advance, is the evaluation still relevant?	<ul style="list-style-type: none"> • Previous evaluations have provided information and recommendations which have been taken into consideration within the programme, demonstrating that evaluations are able to have an influence, e.g. parenting and SLC entitlements.
		<ul style="list-style-type: none"> • <i>The changes to the programme may reduce the applicability of lessons learned as it will be unclear if the outcomes are due to the prior format of the programme or the changes introduced.</i> • <i>Local variations in delivery reduce the influence of any findings, as they may not be applicable to other areas.</i>

<p>Coordination requirements?</p>	<p>How many other donors, government departments, or NGOs need to be or want to be involved? What forms of coordination are possible and/or required?</p>	<ul style="list-style-type: none"> • Stakeholders, such as Local Authorities, are highly invested in the programme, and have previously been involved with evaluations. Therefore future involvement can be expected. • Welsh Government provides guidance and direction to the Local Health Boards. • Many of the coordination requirements fall to the Local Authorities, who are responsible for the various organisations involved in delivering the Flying Start entitlements. <hr/> <ul style="list-style-type: none"> • <i>There is some tension aims between stakeholders, such as those between Local Authorities and Local Health Boards. Currently Local Authorities receive the funding and they commission the Local Health Boards, but the Local Health Boards have expressed a desire to receive funding directly from the Welsh Government.</i>
<p>Utility</p>		
<p>Who wants an evaluation?</p>	<p>Have the primary users been clearly identified? Can they be involved in defining the evaluation? Will they participate in an evaluation process?</p>	<ul style="list-style-type: none"> • Internal Welsh Government staff <ul style="list-style-type: none"> ○ Policy team regularly engage with evaluation activities and are involved in defining evaluation approached • Welsh Assembly members & Cabinet Secretaries/Ministers <ul style="list-style-type: none"> ○ Are not involved in the evaluation process but have final decision making on the approach to be taken and budget approval. • Local Authorities <ul style="list-style-type: none"> ○ May be willing to be involved in defining the evaluation and have been involved in the evaluation process previously. ○ Some have conducted their own local evaluations of the programme
<p>What do stakeholders want to know?</p>	<p>What evaluation questions are of interest to whom? Are these realistic, given the project design and likely data availability? Can they be prioritised? How do people want to see the results used? Is this realistic?</p>	<ul style="list-style-type: none"> • Internal Welsh Government staff and Welsh Assembly Members and Cabinet Secretaries/Ministers want to know: <ul style="list-style-type: none"> ○ The effects of the programme (in the long and short term) ○ If the programme is delivering against the expected outcomes ○ Value for money ○ If the programme can be improved or refined • In addition to the above, internal Welsh Government staff and Local Authorities want to know: <ul style="list-style-type: none"> ○ If specific entitlements are working ○ What combinations are proving most successful? ○ What entitlements are being used? What

		<p>are being used less?</p> <ul style="list-style-type: none"> ○ How the local delivery compares against the national outcomes
		<ul style="list-style-type: none"> • <i>Many of these aims are difficult with the data currently available, as entitlement use of the individuals within Flying Start areas is unknown.</i> • <i>Some Local Authorities collect more accurate and detailed information than other Local Authorities, so a comparison between local and national delivery is only likely to be possible in those areas with more advanced data collection</i>
What sort of evaluation process do stakeholders want?	What designs do stakeholders express interest in? Could these work, given the evaluation of the questions of interest and likely information availability, and resources available?	<ul style="list-style-type: none"> • Currently, many evaluation designs are valued, as long as it is reliable and can stand up to scrutiny. This includes qualitative and quantitative designs. • Qualitative designs are feasible as it requires less data and does not require control groups. This design is valued for its ability to give rich information about the beneficiaries and perceived programme effects.
		<ul style="list-style-type: none"> • <i>Quantitative designs are less practical due to the lack of available data, but are more suited to answering questions on value for money, delivering against outcomes and levels of take up of entitlements.</i>
What ethical issues exist?	Are they known or knowable? Are they likely to be manageable? What constraints will they impose?	<ul style="list-style-type: none"> • Ethical issues create constraints, but it is possible to overcome them through correct procedure and adherence to ethical guidelines and data protection legislation.
		<ul style="list-style-type: none"> • <i>There are a number of ethical issues that arise when evaluating Flying Start;</i> <ul style="list-style-type: none"> ○ <i>Vulnerable groups – many of the beneficiaries are young children or vulnerable adults, e.g. those being exposed to domestic abuse.</i> ○ <i>Data transfer – the large number of delivery organisations makes transfer of sensitive data a regular occurrence.</i> • <i>Data storage – in order to evaluate Flying Start identifiable data will need to be collected and stored</i>
What are the risks?	Will stakeholders be able to manage negative findings? Have previous evaluation experiences prejudiced stakeholder's likely participation?	<ul style="list-style-type: none"> • Stakeholders have responded to previous evaluations in a constructive manner, and have used the evidence to make changes to the programme.
		<ul style="list-style-type: none"> • <i>There has been a conflict between the perceptions of the Local Authorities and the outcomes of previous evaluations. Many stakeholders have a very positive view of the programme and can be</i>

		<p><i>confused when their perceptions are not upheld by evaluation outcomes.</i></p> <ul style="list-style-type: none"> • <i>Welsh Assembly Cabinet Secretaries/Ministers at risk of reputational damage if evaluation outcomes are negative.</i>
--	--	--

- 5.5. The types of outputs which can be expected from an evaluability assessment include the evaluability of a project and the practicality and utility of an evaluation. In terms of this assessment of the evaluability of Flying Start as a project, there are a number of issues relating to the information available and design which make it difficult to evaluate.
- 5.6. However, there are some key positives of the Flying Start programme, in particular the high level of stakeholder (e.g. Flying Start coordinators) buy in and engagement which arguably increases the effectiveness of the programme. In addition the Flying Start programme was designed and based upon a strong evidence base, grounding the programme in evidence which supports the service use with positive outcomes for children. Some of the key issues which adversely affect the ability to evaluate the programme are discussed below.
- 5.7. One of the key areas outlined in the checklist relates to the data and the inability to confirm which Flying Start entitlements have been used on an individual level. Although the data available to confirm the areas for Flying Start and non Flying Start are clear and defined¹¹, the lack of data on individual basis means we are only able to confirm whether children and families are living within Flying Start areas, not whether they have used any of the Flying Start entitlements. This means that the ability to evaluate the impacts of the Flying Start entitlements is not currently possible without additional data collection.
- 5.8. Additionally, there is not an easily identifiable counterfactual to which the treatment group can be compared. The programme was rolled out to those who are in the most disadvantaged areas and so there are no individuals which can be used as counterfactuals, and as such no counterfactual group, as there are no equally disadvantaged areas not included in the programme.
- 5.9. Related to this is the fact that area-based interventions will be provided to those with a range of needs. In the Flying Start areas it is likely that there are some families who do not have a high need for the intervention, and so the programme will have little impact on them. Additionally, there will be those who choose not to engage with the programme. Both of these groups diminish any impacts observed when comparing Flying Start areas against non Flying Start areas.
- 5.10. The checklist also informs the practicality and utility of an evaluation. It can be concluded that, using the information currently available, answering the question 'Does Flying Start work?' is difficult, given the nature of the programme and the lack of available data. More useful questions might be:
- *'How do Flying Start entitlements vary in their effects upon child and family outcomes?'*
 - *'What is it about the Flying Start programme that can lead to beneficial outcomes?'*

¹¹ With the exception of the small outreach areas.

5.11. However, even answering those is a challenge, as we cannot say which entitlements children and families have actually made use of. This means it cannot be said for certain that different outcomes for families are due to engagement with specific entitlements, and even if the outcomes for families are related to Flying Start entitlements at all.

6. Currently Possible Approaches

6.1 Over the last seven years there have been several evaluations which have attempted to evaluate the Flying Start programme. These are detailed in the Table 2.

Table 2. Summary of the methods, advantages and limitations of each of the Welsh Government sponsored Flying Start evaluations and research projects		
Methods	Advantages	Limitations
Qualitative Evaluation of Flying Start (2009) http://gov.wales/docs/caecd/research/091221-qualitative-evaluation-flying-start-en.pdf		
<ul style="list-style-type: none"> In depth interviews with stakeholders, childcare providers, families using Flying Start entitlements and non-user families 	<ul style="list-style-type: none"> Provides rich information about the experiences of those interviewed. highlights unforeseen problems and issues Provides evidence of interactions between entitlements 	<ul style="list-style-type: none"> Does not provide statistical data Unable to provide evidence of impact Cannot be generalised to other families, providers etc. Does not provide information about levels of entitlement use across all users
Interim Evaluation of Flying Start (2010) http://gov.wales/docs/caecd/research/100715-flying-start-interim-evaluation-en.pdf		
<ul style="list-style-type: none"> Online survey completed by each Flying Start Coordinator Area case studies of Flying Start partnerships Thematic case studies relating to specific entitlements, use experiences, programme development and meeting programme expectations 	<ul style="list-style-type: none"> Online survey can give indication of service level outcomes and potential impacts for families Case studies give highly detailed information about specific areas and themes 	<ul style="list-style-type: none"> Does not provide statistical data Unable to provide evidence of impact Cannot be generalised to other families, areas etc. Cannot make conclusions about delivery of expected outcomes
Evaluation of Flying Start: Baseline Survey of Families (2011) http://gov.wales/docs/caecd/research/111214EvalFlyStart-7-20monthsmainen.pdf		
<ul style="list-style-type: none"> Quasi-experimental survey included in-home face to face interviews and self-completion survey of families in Flying Start delivery areas A similar survey administered to families in comparison area. First wave of longitudinal survey 	<ul style="list-style-type: none"> Matched comparison group identified via statistical matching for each outcome indicated measure Provides an estimate for the average treatment effect on the treated via regression and matching 	<ul style="list-style-type: none"> Does not / can not provide information about how and why families access entitlements or any resulting outcomes The impact estimates generated are not completely unbiased Self report method used also

		could be subject to bias
Flying Start Qualitative Research with High Need Families (2013) http://gov.wales/docs/caecd/research/131014-flying-start-qualitative-research-high-need-families-en.pdf		
<ul style="list-style-type: none"> Area case study synthesis report detailing how FS entitlements are being delivered per LA – 60 interviews in 5 FS areas Sample focused on those considered 'high need' 	<ul style="list-style-type: none"> Interviews conducted in 5 FS LAs across Wales to explore experiences in different areas Case study areas chosen to ensure mix of different geographical, demographical characteristics and service delivery models Rich information can be gathered through case studies 	<ul style="list-style-type: none"> Views obtained are not statistically representative of all FS eligible families Impacts are self-reported and reflect changes parents believe to have happened, rather than those that have been objectively measured
National Evaluation of Flying Start: Area Case Study Synthesis Report (2013) http://gov.wales/docs/caecd/research/131128-national-evaluation-flying-start-area-case-study-synthesis-report-en.pdf		
<ul style="list-style-type: none"> Part of series produced by SQW and Ipsos MORI for the national evaluation of Flying Start 22 in depth case studies of Flying Start involving interviews with over 150 stakeholders. Reports on intermediate outcomes for children and families using administrative data 	<ul style="list-style-type: none"> Case studies give highly detailed information about specific areas 	<ul style="list-style-type: none"> Can not decipher impacts from Flying Start solely. Due to timing of report children only just nearing the end of the Foundation Phase, longer term impacts cannot be established Few of the older children would have been able to access the full Flying Start provision
National Evaluation of Flying Start: Impact Report (2013) http://gov.wales/docs/caecd/research/131205-national-evaluation-flying-start-impact-report-en.pdf		
<ul style="list-style-type: none"> Second wave of longitudinal survey, including cognitive assessments of children Delivered survey to 2,116 parents – 1,033 in FS and 1,083 from selected comparison group An intention to treat approach used, meaning families living in areas receiving Flying Start funding were surveyed rather than those who were users of specific Flying Start entitlements. 	<ul style="list-style-type: none"> Matched comparison group identified via statistical matching for each outcome indicated measure Study useful in allowing to build a broad picture of the influence of the programme 	<ul style="list-style-type: none"> Comparison group only an approximation for a counterfactual Does not / can not provide information about how and why families access entitlements or any resulting outcomes More than likely that estimates under-estimate the impact of Flying Start Lack of baseline means it is impossible to know if outcomes are significantly different from pre-intervention levels
Data Linking Demonstration Project: Flying Start (2014) http://gov.wales/docs/caecd/research/2014/140131-data-linking-demonstration-project-flying-start-en.pdf		
<ul style="list-style-type: none"> Experimental approach to identifying children living in Flying Start eligible (FSE) 	<ul style="list-style-type: none"> Can show quantitative impact on a national level for a range of indicators 	<ul style="list-style-type: none"> Cannot provide information on which Flying Start entitlements were used by FSE children

<ul style="list-style-type: none"> addresses and those in the next most deprived (NMD) areas and rest of Wales (RoW) Aimed to demonstrate potential of approach, rather than produce robust outcomes Linked educational and health administrative data to these addresses Compared outcomes of FSE, NMD and RoW children 	<ul style="list-style-type: none"> No need for additional data collection Uses whole population of FSE children Can track individuals beyond involvement with Flying Start 	<ul style="list-style-type: none"> FSE children living in areas that are systematically different to the NMD and RoW Relies of data that was collected for administrative purposes, i.e. not with research in mind
<p>Qualitative Research with Flying Start Families (2016) http://gov.wales/docs/caecd/research/2016/160118-qualitative-research-flying-start-families-wave-1-en.pdf</p>		
<ul style="list-style-type: none"> 216 face to face semi-structured interviews with families in 6 LAs over three years 18 semi-structured interviews with flying start co-ordinators- 3 waves of six interviews annually. Qualitative research with families in non Flying Start areas will take place 2016 - 2017 	<ul style="list-style-type: none"> Provides in-depth evidence to help understand parents' views regarding their experiences of Flying Start The qualitative approach provides insight into families' experiences of FS and how it has influenced their lives Attempts to gather information on what may or may not happen in the absence of the programme 	<ul style="list-style-type: none"> Does not provide statistically reliable data Unable to provide objective evidence of impact Data validity and reliability – the reliance of self-reported data from parents Attribution – interpreting to what extent perceived outcomes and impacts reported by families can be attributed to elements of the Programme

- 6.2. Previous methods of evaluating Flying Start were selected because of the design and roll out of Flying Start, in that the evaluations were commissioned / completed after implementation. This meant that previous attempts at evaluating Flying Start experienced the issues outlined in the evaluability checklist, including the inability to confirm impact due to the lack of available data and the non-existence of a counterfactual or baseline. This excludes some evaluation approaches, such as Randomised Control Trials (RCT).
- 6.3. RCTs are considered by many to be the most robust means of determining causation as it involves randomly allocating the intervention to equally eligible groups. This provides treatment and control groups, and therefore, a counterfactual allowing for comparison of groups; those who have and have not received the intervention.
- 6.4. This approach is not possible (now or in the future) due to implementation already having occurred, with all eligible areas receiving the intervention and control groups not being specified. Without pre-specified control groups and random implementation it is impossible to produce a counter-factual. A baseline of each of these groups is also required. The areas that are not included in the programme are systematically different from those that are.
- 6.5. In the absence of a randomised implementation, phased implementation can serve as an alternative, providing opportunities for quasi experimental designs. In this, all children who are eligible receive the intervention, but some later than others. This has been done to some degree in Flying Start, due to the expansion in 2012, but those who were in the

expansion areas were not included in the original rollout due to living in slightly less disadvantaged areas, and so cannot be considered to be a true counterfactual.

- 6.6. In addition to the range of methods already employed and those not possible, some other approaches could be considered for future evaluations. The evaluations so far have made use of qualitative interviews, focus groups and case studies, quantitative surveys and analysis of existing data, such as those from administrative sources.
- 6.7. When data is collected through a survey or administrative sources, a range of possible statistical approaches can be used to produce an estimate of the impact. Statistical methods are used to understand if these differences are significantly different from what we would expect by chance, i.e. account for random changes in the outcome over time and between groups.
- 6.8. The 2011 Baseline Survey of Families and the 2013 Impact Report employed a survey to collect numerical data that was analysed in this way. At present, a survey is the only way to have any detailed measure of impact, as it is able to collect detailed information about participants. The data currently collected on Flying Start families is not sufficient for a robust analysis of the effect of Flying Start entitlements, only the effects of living in a Flying Start area.
- 6.9. A number of statistical methods are described below, with their relative merits and applicability to Flying Start as it currently exists.

a. Difference-in difference analysis

- 6.10. This method attempts to determine impact by comparing the change observed in the treatment group with that observed in a suitable comparison group. This method assumes that the differences, not related to the intervention, between the treatment and comparison groups remain the same over time. This provides an estimate for what would have happened in the absence of the intervention. This method can compare areas, but not individuals.
- 6.11. This method is not able to give a reliable estimate of the impact of Flying Start, as the differences between Flying Start and non Flying Start areas are considerable, and so it cannot be expected that the differences between them would remain constant (especially given the range of other Government initiatives aimed at supporting disadvantaged areas and households). Also, the lack of a baseline means the change observed in both areas cannot be said to be significantly different from pre-intervention levels.
- 6.12. Also, for this method to be valid, it needs to be expected that the impact of the intervention on the comparison groups would have been the same as the treatment group. Since the Flying Start specifically targets disadvantaged areas, which are unlikely to have received the services offered in the absence of the program, it cannot be expected that the Flying Start programme would impact the rest of Wales in the same way

b. Regression discontinuity analysis

- 6.13. This method of analysis requires a programme to include the whole population but to have continuous eligibility criteria, i.e. a cut off point on a scale, such as age or deprivation index. It then compares the outcomes of those who are only just eligible (i.e. just below the cut off) with those that are only just outside the eligibility criteria. For

example, if a programme was aimed at those who are up to the age of 35, it would compare those in the programme who are 35 against those who are excluded for being 36.

- 6.14. It is assumed that these two groups are similar in their characteristics, and so any difference between them in terms of their outcomes is due to the programme. This creates an estimate for the counterfactual by comparing very similar, but not identical groups. This method could be applied to Flying Start by comparing the outcomes of areas that were only just included in the programme against those that were just outside of it.
- 6.15. However, this method requires a clear cut off point from which the two groups can be derived. While Flying Start was based on the WIMD score, it also was targeted based on the proportion of young children living in that area. It may be possible to produce some specific areas (inside and outside of the programme) that are very similar, but the number of children included in these areas is likely to be small.
- 6.16. One alternative is to compare outcomes of the areas included in the original rollout with the areas which were included in the 2012 expansion. The expansion areas were considered the 'next most disadvantaged' and so could represent the groups that were close to the original eligibility criteria. However, since these areas now are included in the programme, this would only be appropriate using existing data rather than future data collection, i.e. a survey.

c. Matching techniques

- 6.17. Matching techniques rely on using data held on individual programme participants and from those outside the programme. This identifies those included in the programme that have similar characteristics to those outside of the programme. It is assumed that if they have similar characteristics, any differences in their outcomes are due to engagement with the programme.
- 6.18. Propensity score matching is a technique which attempts to determine the ability of each of these characteristics to predict the likelihood of each child either being in the treatment or control group. There will be some outside of the treatment group that have characteristics that would suggest they are more likely to be included than excluded. For example, for Flying Start the WIMD and HMRC and DWP data was used to determine eligibility, so those in Flying Start areas are assumed to be more disadvantaged than those outside of these areas. However, at a household level, they are going to be some who are equally, if not more, disadvantaged outside of Flying Start areas than inside. It is these households that can be matched and compared.
- 6.19. This approach can show that there is a difference between groups, but does not give any indication the causes of the differences. While it is assumed that the difference is related to the programme, the strength of the matching relies on having a range of data items for each programme participant and the same data for those outside of the programme.
- 6.20. A large scale survey, which includes both families in Flying Start areas and those in other areas, could collect the necessary data for application of this method. Both the 2011 Baseline Survey of Families and the 2013 Impact Report made use of this technique. The survey conducted included roughly 3,500 families (half of which were eligible for Flying Start entitlements) and attempted to match families based on size, child age, parental

education, housing, lone-parent status and other socio-demographic variables. Any future surveys should try to ensure that the families in non Flying Start areas are as similar to those in Flying Start areas as possible, to ensure a higher level of success when attempting to match between areas.

- 6.21. It is also possible to attempt to apply matching techniques to existing administrative data, such as the National Pupil Database. This would require the ability to identify those children who had received Flying Start services, and then match them with those that had not. However, there are two issues with this;
- a) The matching variables are restricted to what is held in that administrative dataset
 - b) There is not enough data to know which children have made use of Flying Start entitlements (only if they are living in a Flying Start area while in school).

d. Mediation/moderation analysis

- 6.22. There are two possible ways in which a programme's effect may change in different circumstances. The first is mediation, which describes how the effect of the programme on the outcome may occur through a third variable, i.e. the programme has an indirect effect. For example, it may be predicted that Flying Start has a positive impact on children's literacy, but this occurs by improving the parent-child relationships. If the relationships do not improve, then there is no impact on literacy.
- 6.23. The other means by which the effect is altered is moderation, where the effect of the programme on the outcome is altered by the presence/absence of a third variable. In this case, Flying Start has a direct effect on improving literacy, but if parent-child relationships are good, then the effect is stronger.
- 6.24. This analysis usually involves testing a wide range of relationships between variables and attempting to understand the circumstances in which the effects occur and when they do not. This analysis is only possible when there is a lot of data for each individual, so that the relationships between variables can be tested. It is currently possible to test the relationship between living in a Flying Start area and a range of outcomes. However, without individual level data, it is not possible to test the interactions between receiving Flying Start entitlements and other demographic variables and also the relationships between different Flying Start entitlements.
- 6.25. With these methods considered, there are a range of research questions that can currently be answered. This includes those that have already been addressed in previous evaluations, and those that have not which are potentially possible. Many of these have already been looked at as part of previous evaluations, and so without changes to the programme and the data collected, future evaluations will only be able to replicate previous findings.
- Do children living in Flying Start areas have different outcomes to those in non-Flying Start areas?
 - Do children living in Flying Start areas after implementation of the programme have different outcomes to those living in the same areas before implementation?
 - Do families living in Flying Start area have positive perceptions of the programme and the entitlements provided, in relation of experiences and perceived impact?
 - What specific entitlements do families/parents/children feel is having the most impact?

- What barriers prevent families/parents/children living in the Flying Start areas from engaging with the Flying Start programme?
- Do stakeholders (e.g. Local Authorities) perceive the programme to having positive effects?
- What additional services would families/parents/children want to see provided in Flying Start areas?

- 6.26. In addition to those research questions that can currently be answered, there is a range which are currently impossible to answer without changes to the available data, changes in the way in which the programme is delivered or investment in a large-scale survey.
- What is the level of take up of Flying Start entitlements by individual families?
 - What level of engagement with Flying Start entitlements leads to different outcomes?
 - What are the interactions between the different Flying Start entitlements?
 - What are the outcomes of families/parents/children living in Flying Start areas that do not make use of the entitlements, compared to those that do?
 - What are the longer term impacts of Flying Start? I.e. track those who have used Flying Start entitlements through to adulthood.
 - What are the demographics of those with higher or lower levels of engagement with Flying Start entitlements?
 - Do different individual characteristics lead to different outcomes for those who engage with Flying Start entitlements? E.g. comparison of males and females with similar levels of engagement.

7. Potential Future Approaches

- 7.1. The evaluability checklist (section 5) clearly identifies three main challenges for evaluating Flying Start: lack of data about engagement with specific Flying Start entitlements at the individual level, lack of counterfactual and area-based delivery, including differences in the delivery between areas. Without substantial changes to the programme or monitoring and data collection, there is nothing that can be done to address the issues presented by the latter two challenges. However, it is possible to start collecting additional data to support a more robust evaluation.
- 7.2. In order to be able to answer any questions regarding the impact of the specific Flying Start entitlements, it is necessary to know the level of engagement with Flying Start entitlements at the individual level. Ideally this data would include all the entitlements which the families/parents/children in Flying Start areas had made use of. It would also contain a number of demographic variables, such as the age, family size, ethnicity, children's gender, parents' employment status and qualifications and household size.
- 7.3. This data would be enhanced by being able to link it to other datasets collected by other services. For example, linking the data to that collected by schools would allow for a detailed analysis of whether Flying Start entitlements lead to improved educational outcomes, and if so, what specific entitlements are more impactful.
- 7.4. It would also be useful to gather information to establish the use of other interventions for which the family/parent/child has engaged with, such as Communities First or Families First. This would allow for better understanding of whether the Flying Start entitlements

alone are responsible for any impacts/ changes observed or whether changes are more likely to be found where there has been engagement with other interventions in addition to Flying Start.

- 7.5. This additional data would allow for statistical analysis methods to be applied to a much greater degree. There would be the opportunity to look at comparisons based on the actual entitlements used by children, rather than comparison on an area-level basis, for all children eligible for the programme. This relates to the aims of Flying Start more accurately, as Flying Start was designed to provide holistic, tailored support to families in Flying Start areas. The issues with area-based comparisons is that there is a range of take-up and need within an area leading to a dilution effect, this would not be an issue if the actual levels of engagement were known.
- 7.6. While individual data would allow for a range of additional research questions to be answered, the collection of it would not be a simple undertaking. It is highly likely that it would require substantial resources from both the Welsh Government and Local Authorities. Another issue with using this data for evaluation purposes is that there is still a lack of a comparison group. While it would be useful to have additional data for participants of the Flying Start programme, similar data items will be required from those living in non Flying Start areas. For an accurate comparison, the same items will need to be collected for both Flying Start and a sample of non Flying Start families/parents/children.
- 7.7. There are two possible routes to collecting this data: a survey of participants or developing a new data collection system for all participants of the programme. While a sample survey has been used in previous evaluations (such as the 2013 Impact Report), it is still worth considering for future evaluations. However, it needs to be compared against alternatives in order to determine the most appropriate approach.

1. Sample Survey

Gains

- Understanding of Flying engagement at a national level
- Snap-shot of numbers of participants for each of the entitlements
- Potential to determine interactions between entitlements with a large enough sample
- Longitudinal survey could track development and outcomes over time
- Cross-sectional survey able to compare outcomes for groups at different stages in the programme
- On-off data collection which does not require ongoing costs

Risks

- Need sample of sufficient size to provide reliable analysis
- Suitable comparison group needs to be identified
- Would need to be repeated to assess new different or future cohorts
- Poor response rate and/or attrition of participants¹² (if longitudinal)
- Possible response bias from those most willing to engage in Flying Start areas

¹² Longitudinal surveys attempt to survey the same participants over a range of time points, often several years. However, some of these participants may choose to drop out of the research before the survey is concluded. The attrition rate is the proportion of those participants that do not complete the whole survey.

Resource Implications

- High cost for a large-scale survey, potentially representing a significant proportion of the total Flying Start budget
- Staff required to manage and oversee the survey and/or manage the contractor
- Lengthy process from survey development to obtaining results

2. Individual level data collection

Gains

- Detailed data on engagement with all entitlements of the Flying Start programme
- Up-to-date and ongoing information about exact numbers of participants for each of the entitlements
- Ability to determine interactions between entitlements
- Large dataset allowing for robust analysis of the whole population of children living in Flying Start areas
- Linkable to other datasets, such as National Pupil Database, which would allow for analysis of wider outcomes
- Can track Flying Start participants through the programme and compare those at different stages of the programme
- Able to assess potential effects of Flying Start on an individual basis

Risks

- Data Protection and sensitivity of items collected. Consent to share the data would need to be gained
- Data management and storage tools need to be considered and developed
- The process of collation would need to be consistent across Local Authorities
- Validation of the data and practicality of the actual collection
- Data collected by non-researchers, possibly leading to systematic biases
- Some local areas may lack the systems necessary to collect individual level data, at least in the short term.

Resource Implications

- Data management systems may need to be developed by either Welsh Government or local authorities, likely by a third party contractor which could be at a high cost
- Staff with specialist skills to manage the collection locally and centrally
- Data collection would need to be managed for the duration of the Flying Start programme, i.e. long term resource commitment
- Time to collect and match the data in the Local Authorities

7.8. Therefore, it may be a challenge to implement the necessary systems for individual data collection, and use of a sample survey may not provide the necessary data. However, there are substantial gains to be made and risks if the data continues to not be collected. Flying Start is a large investment and without knowing the true effects of the programme, it is not known if Flying Start is good value for money. In order to determine where

possible improvements can be made, there needs to be full comprehension of the effects of the entitlements gained through gathering individual level data.

8. Conclusions

- 8.1. The aim of this report was to determine if:
 - It is plausible to expect impacts
 - It is feasible to measure impacts
 - Whether an impact assessment would be useful
- 8.2. Previous evaluations of the Flying Start programme have all adopted methods constrained by the design and roll-out of the programme, in that these were commissioned after its implementation. The previous evaluations have not been able to determine the effects of the Flying Start programme because of data limitations and the programme design.
- 8.3. The data presently available only allows for analysis of those eligible for Flying Start rather than for recipients. In addition, there is no group of families which can be used as a counterfactual to compare to those who have received the Flying Start programme, as the programme has been rolled out in the most disadvantaged areas of Wales, of which there is no equivalent to use as a counterfactual.
- 8.4. The programme was developed (and continues to be adapted) in line with the best available evidence, which suggests it is plausible to expect beneficial effects for the children. However, it is unclear what these impacts necessarily should be, due to the unavailability of data of those using the Flying Start entitlements. Therefore it is difficult to measure impacts. Additional data collection is required, either collected via a survey or by improving monitoring data. For an outcome evaluation to be useful, it will need to be able to identify what entitlements the family/parent/child has made use of, and what the outcomes from these are. There is currently not enough available data, of the right kind, to determine this.
- 8.5. The lack of a counterfactual cannot be so easily overcome, because Flying Start was implemented in the most disadvantaged areas in Wales, meaning that any prospective comparison areas are inherently dissimilar to some degree. However, collection of additional data for Flying Start and non-Flying Start areas should allow for more accurate and detailed comparison groups.
- 8.6. In summary, some outcomes can be determined now, but these are limited:
 - It is possible to reach findings about:
 - Parents' and stakeholders' attitudes and perceptions of the Flying Start programme and its entitlements
 - The differences in some of the things Flying Start is intended to influence, for children living in Flying Start areas compared to other areas
 - The changes in some of the things Flying Start is intended to influence for children living in Flying Start areas before and after its implementation
 - It is not currently possible to assess:
 - The impact of Flying Start entitlements on children's long term development

- The specific effects of Flying Start entitlements, or their interaction with each other
- The actual take-up of entitlements by individual families, and groups that have higher levels of take-up

Recommendations

- 8.7. This report makes the following recommendations for any future evaluations of the programme in its current form:
- To compare the intended outcomes between Flying Start and non-Flying Start areas using statistical matching for groups of children in each area who have similar characteristics
 - To continue to use qualitative research to understand the perceptions of families living in Flying Start areas
 - To explore options for data linking, using administrative data to potentially identify individual level outcomes for children living in Flying Start areas
 - To identify a sample of families/parents/children who have engaged with Flying Start entitlements and track them longitudinally

The possible improvements to the programme that could lead to more robust evaluations:

- Develop new systems for collecting individual level data about engagement with Flying Start entitlements, which could include either additional monitoring and/or a large scale survey
- Link individual level data collection to data collected by other services, such as schools' educational data
- Identify where families/parents/children are receiving additional services or interventions, and link this to the individual level data from Flying Start.

If these improvements were to be made, the following activities may then be possible:

- Using statistical matching, but at an individual level to determine the effects of specific Flying Start entitlements.
- Analysing the individual level data to test whether individual characteristics, or wider contextual factors, affect take-up and impact of services.
- Using individual level data for selecting samples for further qualitative research and/or more in-depth surveys, case studies and longitudinal research.
- Identifying whether additional services or interventions (other than Flying Start) are having an additional effect on those in Flying Start areas.