EARLY INTERVENTION FOUNDATION

Evaluating early help

A guide to evaluation of complex local early help systems

March 2019

Sarah Taylor, Elaine Drayton and Tom McBride

Acknowledgments

We are very grateful to the local authorities named in this report.

We are also grateful to Kirsten Asmussen, Lucy Brims, Jess Ford and Inês Pote (EIF), Josie Dickerson (Better Start Bradford Innovation Hub), Jacob Botham, Julian Cox, David Ottiwell and Nigel Waddington (Greater Manchester Combined Authority), Ian Dean and Louise Smith (Local Government Association), the Department for Education, the Ministry of Housing, Communities and Local Government, and Dione Hills (Tavistock Institute). All analysis and findings presented here reflect the views of the authors only.

About EIF

The Early Intervention Foundation (EIF) is an independent charity established in 2013 to champion and support the use of effective early intervention to improve the lives of children and young people at risk of experiencing poor outcomes.

Effective early intervention works to prevent problems occurring, or to tackle them head-on when they do, before problems get worse. It also helps to foster a whole set of personal strengths and skills that prepare a child for adult life.

EIF is a research charity, focused on promoting and enabling an evidence-based approach to early intervention. Our work focuses on the developmental issues that can arise during a child's life, from birth to the age of 18, including their physical, cognitive, behavioural and social and emotional development. As a result, our work covers a wide range of policy and service areas, including health, education, families and policing.

EIF IS PROUD TO BE A MEMBER OF
THE WHAT WORKS NETWORK

What Works Network

Early Intervention Foundation
10 Salamanca Place
London SE1 7HB

W: www.EIF.org.uk
E: info@eif.org.uk
T: @TheEIFoundation
P: +44 (0)20 3542 2481

The aim of this report is to support policymakers, practitioners and commissioners to make informed choices. We have reviewed data from authoritative sources but this analysis must be seen as a supplement to, rather than a substitute for, professional judgment. The What Works Network is not responsible for, and cannot guarantee the accuracy of, any analysis produced or cited herein.

EIF is a registered charity (1152605) and a company limited by guarantee (8066785).

Download
This document is available to download as a free PDF at:
https://www.eif.org.uk/resource/evaluating-early-help-a-guide-to-evaluation-of-complex-local-early-help-systems

Foreword

Nationally, there is ongoing debate about the extent to which early help makes a difference to children's outcomes.

This is an important question, but one that currently we cannot answer robustly enough. As this report finds, there is a lack of good-quality evidence about the impact of local early help systems on outcomes for children and families. There are a range of early intervention programmes which have been shown to improve outcomes for the children who receive them. However, during the course of this study and in the wider work of EIF, we have not found evaluations of the impact of early help offers that we would consider to be robust. In many ways, this is not surprising. Generating good-quality evidence of impact, particularly of systems involving a range of agencies and services, is difficult. It takes time, and requires capacity, resources and capability, all of which are in short supply in local services.

But this lack of evidence matters. It matters locally because it makes it difficult to know which approaches are most promising, or which features of integrated early help models might make the most difference. Time and resources may be wasted constantly adapting and re-designing services rather than evaluating, learning and improving based on solid learning and evidence.

It also matters because this lack of data on the impact of early help services makes it difficult to make the case for investment in early help both locally and nationally.

An absence of evidence doesn't mean something *doesn't work*. It means *we don't know*.

This is not a good place to be. As local authorities try to find savings, there is a real danger that early help services are cut, not because they don't work, but because we haven't been able to test whether they do.

We need to build capacity to use and generate evidence in children's services. This means equiping those delivering early help with the skills and resources required to measure the impact of their services and, in turn, to generate good-quality evidence. Routine use of this evidence should become part of 'business as usual', informing decisions in local authorities and other commissioning bodies. Improving management information and tightening up objectives and measurement should be something all areas aim to do as part of their continuous improvement plans.

This guide seeks to make an initial contribution to improving the evaluation of early help by setting out some principles that could help places to make progress on evaluating local services. It is aimed at managers and commissioners, and describes how they can make services both easier to evaluate and more evidence-based.

Local authorities have a vital role to play in taking this forward, but they cannot do it alone, particularly in the current financial climate. Action is also needed centrally, at a national level, to put in place the support local areas need to apply these principles, so that they can develop and implement systems to robustly evaluate their early help arrangements. This is a priority if we are to make some much-needed progress in filling this gap in the evidence base for early intervention in the UK.

Donna Molloy
Director of Policy and Practice, Early Intervention Foundation
March 2019

Summary

About evaluation

Many types of evidence make a useful contribution to understanding families and the services they receive. But only **impact evaluations** can verify whether a service works, and tell us how much difference it makes, relative to not receiving the service at all. Impact evaluations need quantitative data, though they can also involve qualitative methods.

The best impact evaluations, which can draw causal links between services and family outcomes, involve randomising which families receive or don't receive services. Due to its complexity, early help is difficult – but not impossible – to evaluate in this way.

About this guide

This guide is about how to evaluate the complex systems that make up local early help offers, which may involve many partners and multiple programmes and services. This is not easy, but the examples in this guide show that it is possible. This is an emerging area, which no local authority has yet fully worked out. The six principles set out in this guide can help managers and commissioners to find out what difference they are making for families, as part of the smooth running of evidence-based, continuously improving services.

Principle 1: Know where you are starting from

You need a good understanding of local families if you are to evaluate the services they receive. Involve your partners so that you can set out the whole system and understand how all local activities contribute to outcomes. Your **needs assessment** should go beyond describing needs at a point in time to consider pathways of needs – how needs change over time for different families, strengths (such as strong relationships), and patterns of service use. You also need to understand the inputs, activities, outputs and outcomes of your local offer, and the assumptions and external factors on which these depend, as part of a **logic model**. One way to build your understanding is through **system mapping** of relationships and pathways. Consider making changes to data systems in order to track families and capture links between services.

Principle 2: Prioritise outcomes to evaluate

Set the scope of your evaluation by defining where the boundaries of your early help system lie. You and your partners in the early help system should only evaluate **the outcomes for which you are accountable**. Many local areas have broad aims, such as better outcomes, improved user experience, and cost savings. **Clearly defined aims** give you something concrete to evaluate. Your aims should be realistic, and you should limit yourself to a **manageable number of aims**.

Principle 3: Embed evaluation in commissioning and practice

See evaluation as part of commissioning, not as a one-off activity. Invest in evidence capacity as part of continuing professional development: **upskill staff** on how to understand, interpret, and gather data for evaluations. Demonstrate leadership by using evidence, and **grow an evidence culture** across the system. Prioritise **investment in evaluation** as part of commissioning. One rule of thumb is to assign at least 10% of a programme's budget to evaluation.

Principle 4: Use high-quality measures

Exploit existing sources of routinely collected data on outcomes, and only gather data yourself where these sources are not available. If you do gather your own data, use measures that have been carefully developed by experts. **Measures should be valid and reliable**, relevant and responsive; reasonable in terms of length, simplicity and accessibility; allow you to **track distance travelled** or progress; and broad enough for use with a range of families. Consider upgrading data systems and changing working practices to introduce a shared ID, such as a unique pupil number or NHS number, for all partners to use.

Principle 5: Make comparisons

Collect data from families both **before and immediately after** they receive support from early help services. Then **follow up** to see whether changes are sustained. Make strenuous efforts to follow up with as many families as possible: even if they drop out of receiving services, they can still contribute data to the evaluation. Carry out **statistical testing** on whether the difference between the 'before' and 'after' data is significant, or if it is likely to have been due to chance. Going further, recruit or statistically construct an appropriate comparison ('control') group who do not receive early help, ideally through randomisation.

Principle 6: Follow through

Be **open and transparent** by publishing your findings, and acknowledge the limits of the data and analysis. If your methods do not allow you to make causal claims, do not make them. Provide enough detail to allow someone elsewhere to replicate your findings. Plan to develop and improve the quality of your evidence and evaluations over time, to increase the robustness of your findings. **Act on your findings**, implementing the recommendations made in your evaluation reports. The large number of influencing factors and massive complexity within an early help system means that large sample sizes are needed to distinguish real improvements from random changes. **Be patient**: wait for the number of cases to build up and for long-term outcomes to emerge.

Contents

Foreword 3

Summary 4

Introduction 8
- How is early help evaluated at present? 8
- Why is it important to evaluate early help? 8
- Opportunities and challenges in evaluating early help 9
- Scope of this report 10
- What is early help? 10

About evaluation 13
- What is impact evaluation? 13
- What is *not* impact evaluation? 14
- Combining impact evaluations with other evidence 15

Why is early help difficult to evaluate? 17

About this guide 19

Principle 1: Know where you are starting from 20
- Why knowing where you are starting from is important 20
- Assessing your area's needs 20
- Thinking about needs in the round 23
- Logic mapping: how and why you believe you are helping families 25
- Map your system 26

Principle 2: Prioritise outcomes to evaluate 29
- Why prioritising outcomes to evaluate is important 29
- Set the scope of your local early help offer 29
- Define success: what does good look like? 30

Principle 3: Embed evaluation in commissioning and practice 32
- Why embedding evaluation in commissioning and practice is important 32
- Evaluation is part of commissioning 32
- Invest in evaluation 33
- Build a culture of evidence-based practice 34

Principle 4: Use high-quality measures 35
- Why using high-quality measures is important 35
- Where available, use existing high-quality measures 35
- Where no high-quality measures are available, identify valid and reliable measures 36
- Is it a measure of improvement? 37
- Don't reinvent the wheel 38
- Measure outcomes that you are accountable for influencing 39
- Work with families to measure distance travelled towards outcomes 40
- Minimise bias 40
- Sources of validated measures 40
- Consider the population norm for the measure 41
- Measure the unexpected 42
- How to pick a measurement tool 42
- Share unique identifiers across systems and partners 43

Principle 5: Make comparisons **46**
Why making comparisons is important 46
What it means to make a comparison 46
Track change over time 47
Compare 'with and without': families who receive early help versus those who do not 50

Principle 6: Follow through **53**
Why following through on your evaluation is important 53
Be honest and transparent 54
Plan in advance 55
Manage expectations 55
Horses for courses: who to report to 55
What to report 55
Framing recommendations 56
Keep going 56
Be patient 57
Do you have enough families to evaluate? 57

Annex **59**
Questions for councillors to ask council officers about local evidence and evaluation 59
Further reading 61
Examples of evaluation designs and techniques 61
What good evaluation looks like for the Early Intervention Foundation 62

Introduction

Early help services aim to help families at an early stage of problems arising. They are linked local systems of casework and other non-statutory interventions which provide support to vulnerable children, parents and families who are at risk of escalating needs and poor long-term outcomes. Evaluating these early help services is important if we are to understand whether they are delivering as expected and improving outcomes for the populations they serve. Local areas want to know that their early help offer is helping families, but testing this robustly is a complex task, and many areas need help to achieve this.

The Early Intervention Foundation (EIF) is an independent charity and part of the government network of 'What Works' centres. Our mission is to ensure that effective early intervention is available and is used to improve the lives of children and young people at risk of poor outcomes.

Improving the evidence base for early intervention in the UK is vital to our mission. The strongest evidence for early help is currently taken from evaluations of individual early intervention programmes. But there is a lot of early help activity that has not yet been evaluated at all. And the evidence is less developed about what kinds of wider activities or early help systems, with multiple components, are most effective.

One of the most common requests we receive at EIF is for support and advice on evaluation. We aim to provide at least some of that through this guide. We set out practical tools for overcoming the barriers to carrying out robust evaluations of the complex systems of services which make up early help, because growing capacity and capability locally to evaluate these services is a vital part of our work to improve the evidence base for early intervention in the UK.

How is early help evaluated at present?

Some local authorities are already evaluating their early help services. Many more would like to but are struggling with how to do this. People are doing what they can, but there are inherent challenges in evaluating complex and multifaceted systems, and a lack of capacity locally. Questions about impact, attribution and cost savings are not straightforward to answer, and the necessary technical expertise, capacity and resources can be difficult to find.

As a result, there is a striking lack of data currently about the impact of early help, and real scope to improve the quality of evaluation of local early help systems. We highlight examples of good practice in this guide, but we are not aware of any evaluations of local early help systems which are robust enough for judgments on whether the services are the cause of improved outcomes for children and families. Although some local authorities have attempted to evaluate their early help arrangements, often this has focused on evaluating processes or people's movements through services rather than impact for children and families.

Why is it important to evaluate early help?

The lack of robust evaluation of early help systems means we cannot be sure what difference these systems are making for families. This makes it more difficult to make the case for investment in early help locally and nationally. It also makes it difficult to know what

approaches are most promising, or which features of integrated models might make the most difference. Commissioners and service managers may be working in the dark. Time and resources may be being wasted repeatedly innovating rather than testing, learning and building on previous initiatives.

Local authorities are accountable to local electorates for the value for money of local spending on early help. Impact evaluations provide key evidence to inform judgments on the effectiveness of services. Taking evidence-based decisions, informed by evaluation, reduces the risk of wasting money on services that will not improve the intended outcomes for vulnerable children and families.

Many early help professionals who support families are certain that their work makes a positive difference. They may have witnessed compelling testimony from parents, saying it was their help that made all the difference. However, while service user testimony and professional judgment are important, they do not in themselves prove impact.

Impact evaluations are different to other kinds of evidence because they can confirm causal links between services and outcomes: did the service cause a family's outcomes to improve? Good-quality impact evaluations measure performance against a comparison (a counterfactual) to estimate how much good services do, over and above what would have happened to families in the absence of early help. They are able to judge whether families' outcomes would have improved anyway.

Opportunities and challenges in evaluating early help

Some features of early help actually support local evaluation.

- Its **non-statutory nature** means services are discretionary. This offers the possibility of using control groups, who receive services only after a delay or not at all.[1] Control groups make evaluations more robust.
- A **culture of evidence-gathering** already exists to some extent in the sector, as seen in the widespread use of assessments. This is a good foundation for developing a culture of evaluation, and for introducing more robust ways of demonstrating impact.
- **The sector is data-rich,** thanks to the returns to central government that are required by the Troubled Families Programme. Some of this data can be exploited as part of local evaluations.[2]

Other features of early help make local evaluations challenging.

- **Breadth of ambition**, in that early help systems aim to help families with a wide range of problems, making it more difficult to pin down what outcomes an evaluation should focus on.
- **Family life is complex**, and early help recipients may be receiving support from a range of services at the same time, making it difficult to identify the 'active ingredients'.
- **Systems change over time**, making it difficult to determine which features of the changing system have influenced family outcomes.
- **Lack of capability**, with gaps in data, lack of funding for evaluation, and variable levels of skill and confidence in generating evidence among those involved in delivering early help.
- **Separate data systems used by different agencies** do not facilitate evaluation, as software often does not allow linking between different parts of an early help system.

1 For a discussion of the ethical considerations of evaluation involving non-participants, see Principle 5: Make comparisons.

2 See Greater Manchester Combined Authority example on page 49.

Scope of this report

Various guides are available on how to evaluate. These tend to cover evaluation practice in general, or how to evaluate individual programmes, as in our guide, *10 steps for evaluation success*.[3] The gap this guide aims to fill is on how to carry out, or intelligently commission, impact evaluation on local systems of early help. Our emphasis is on how to make services easier to evaluate, through improvements to a business-as-usual approach.

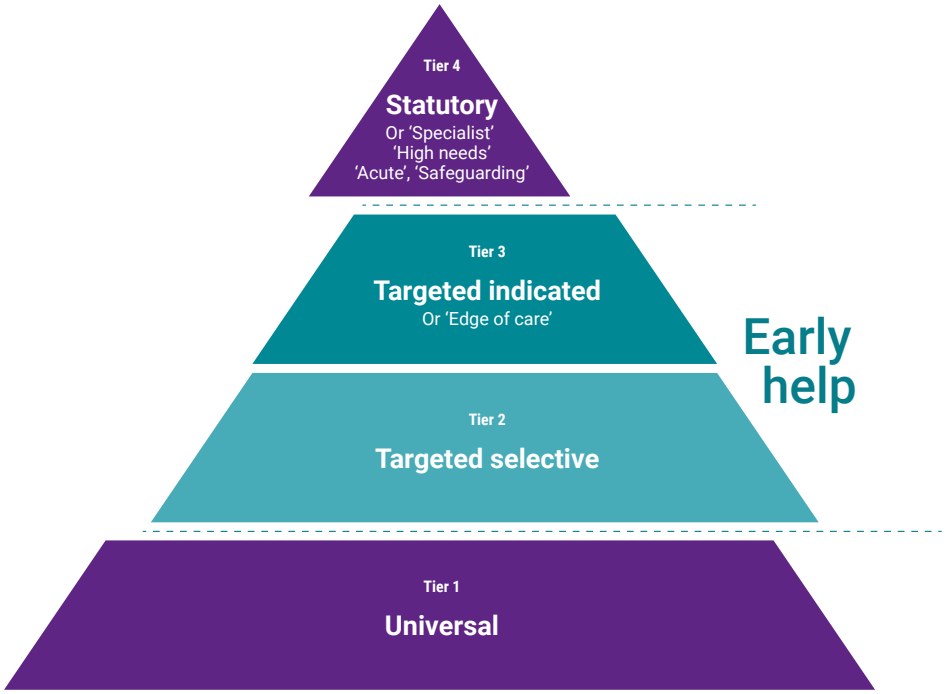
In our view, elected members have a vital role to play. They can approve capacity and resources for evaluation and ask officers for this evidence and scrutinise its quality; we include a list of questions which members might ask in the annex.

What is early help?

The term 'early help' came to prominence in Professor Eileen Munro's review of child protection (2011).[4] Early help services are non-statutory and can be universal or targeted. They often cover services that fall between universal and statutory, or tier 2 and tier 3 services as shown in figure 1 below. This often includes targeted support from universal services such as schools or the NHS, and coordinated family support services. Returns to the Department for Education on early help spending cover areas such as Sure Start children's centres, the Troubled Families Programme, and support for young people, such as youth groups or teenage pregnancy services. Resourcing constraints in recent years mean early help in some areas has been restricted to families with higher levels of need. But government guidance is clear that local areas should have effective, evidence-based early help services in place.[5]

FIGURE 1

The term 'early help' is commonly used for services that fall between universal and statutory



Source: EIF. Note: Some local areas also consider universal, tier 1 services to be part of early help; others focus their early help within only one of the tiers.

[3] See: https://www.eif.org.uk/resource/10-steps-for-evaluation-success

[4] See: https://www.gov.uk/government/publications/munro-review-of-child-protection-final-report-a-child-centred-system

[5] Department for Education (2018) *Working Together to Safeguard Children*, page 15: https://www.gov.uk/government/publications/working-together-to-safeguard-children--2

Common features of early help

Early help systems vary, but often have features in common, as illustrated in figure 2.

FIGURE 2

Common features of coordinated multi-agency early help

Source: EIF

The services and programmes covered by early help can be delivered in the family home, or involve family members visiting children's centres or other public services such as schools. They often include family and parenting support, supporting family relationships, supporting aspects of children's development such as building social and emotional skills, tackling behavioural issues or supporting children's learning, and practical help with issues such as employment, housing or debt. These are delivered by a range of professionals, and sometimes volunteers, from family support workers, to health visitors, to police officers.

Differences between early help offers

Other features vary from place to place, including whether early help is provided to families using universal services (tier 1), to families with low-to-moderate needs on the edge of universal services (tier 2), and/or to families with higher levels of need on the cusp of needing statutory help (tier 3). Other differences include the thresholds for these tiers, and the types of support offered at each level.

This variety makes it more important to evaluate the effectiveness of your particular local early help offer, as well as referring to evidence gathered elsewhere. Evidence that something has worked in the past is important, but does not by itself guarantee similar results in the future or in a different location.

What does an early help system look like?

Wandsworth example

The London Borough of Wandsworth is still developing its new early help offer. It involves a universal front door (or single point of contact) for concerns about children, young people and families below the threshold of social care. Where a child is identified as needing targeted support or help that cannot be met through universal services, an early help assessment is carried out. This is used to develop a Signs of Safety and Wellbeing plan. Practitioners work in partnership with the child and family to form a Team Around the Family, who work together to deliver the plan and to build resilience in the family. They track the progress of families using the Mosaic system.

The aims of Wandsworth's early help are to:

- proactively reach families who will benefit the most
- holistically cover all aspects of wellbeing
- provide the right intervention in the right place at the right time
- work in new, creative and empowering ways
- value and support families, staff and partners to improve the offer
- provide services are based on evidence about what works.[6]

[6] For more, see: http://www.eif.org.uk/case-study/thrive-wandsworth/

About evaluation

What is impact evaluation?

Impact evaluation asks what difference something makes. The focus is usually a defined programme or intervention, rather than a broad service or system. Impact evaluations seek to reach a judgment on whether a programme has had an impact or not by testing what changes have occurred, and the extent to which they can be attributed to the programme. In short, has there been a change, and did the programme cause that change?

Robust impact evaluations isolate the effect of a programme from all other influences on families. To test how responsible a programme was for an outcome, evaluators estimate statistically what would have happened in the absence of the programme. This provides the comparison to judge performance against. Numbers – quantitative data – are therefore needed for robust impact evaluations. Words – qualitative data – cannot by themselves answer the question of whether a programme works, although they might explain why a programme does or does not work. The best evaluations combine both types of data to reach a comprehensive judgment on a programme's effectiveness.

Strong impact evaluations which make valid comparisons between groups which do receive programmes and those who do not can confidently attribute improvements to programmes themselves, as opposed to other factors influencing families.

Good impact evaluations analyse data collected at two time points – before and after the programme – and for two different groups, with and without the programme. This approach is only legitimate if you can prove you are comparing apples with apples: the two groups (known as the intervention group and the control group) should be compared to ensure they are broadly similar in all important respects, including the outcome of interest, before this design can be considered robust. Evaluations without both data from 'before and after' and 'with and without' fall short of being able to say what works.

Methods used in impact evaluations

There is no off-the-shelf answer to the question of how to carry out an evaluation. The best design and methods will depend on the aims of your evaluation, and what is feasible and proportionate in the circumstances. It is good practice to frame your evaluation questions first, and only then to select the methods.

Methods vary, but impact evaluations can involve randomised control trials (RCTs), quasi-experimental designs (QEDs), or exploiting natural experiments.[7] We do not recommend any particular methods or designs; however, all the highest-rated programmes included in the EIF Guidebook have evidence from randomised control trials.[8]

All these methods require developing or buying in technical expertise. However, this guide describes how to lay the foundations for these more ambitious and demanding methods.[9]

---

7 For definitions and descriptions of these methods, see the annex.

8 See: https://guidebook.eif.org.uk/. The EIF Guidebook is a free online database of information on early intervention programmes for which we have rated the strength of evidence.

9 See also our 2019 report, *10 steps for evaluation success*: https://www.eif.org.uk/resource/10-steps-for-evaluation-success

Questions that impact evaluations can answer

Through impact evaluation, local areas can answer questions about the value added by programmes. Here are some questions that only impact evaluations can answer:

- Is life better for families as a result of what we do?
- Are we ensuring that problems do not escalate?
- What is the impact on higher-level statutory services?
- Are there reductions in the need for specialist services?[10]

For example, an impact evaluation forms part of the national evaluation of the Troubled Families Programme. It asks how much impact the programme has on families, across a range of outcome measures that the programme aims to improve.

What is *not* impact evaluation?

Other types of evidence are needed to answer different types of questions. These types of evidence form part of the broader evidence base for an early help system.

- **Needs assessments:** It is important to understand families' needs and characteristics at population level, by analysing published data sources or carrying out primary research in the local area. In Principle 1 we describe needs assessment as a foundation for evaluation.
- **Monitoring:** Tracking inputs, activities, outputs and outcomes, such as the number of assessments carried out or the types of people helped by a service, gathered as part of business-as-usual or through periodic research, may be formally reported to councillors or published, or used to inform internal operations.
- **Measuring distance travelled:** Descriptions of the distance families travel towards outcomes, or of the intermediate outcomes they have achieved, are not evaluations.[11]
- **Service user feedback:** Feedback, including complaints and compliments, and qualitative data on perceptions of a service, should be central to quality assurance and service improvement, so that services are user-focused and children's voices are taken into account.
- **Case reviews:** Reviews of cases where opportunities were missed to achieve a better outcome for family members are essential for learning and improvement.
- **Audits and inspection reports:** These can also be invaluable sources of evidence, tending to focus on compliance with process.

Evaluators can use these forms of evidence. They can be suggestive, but not conclusive. They are not in themselves impact evaluations, and should not be used to make claims about effectiveness.

Questions that impact evaluations cannot answer

Local areas should draw on research, monitoring data, process evaluations and other evidence to answer 'who', 'why', 'when', 'where', and 'how' questions. By contrast, impact evaluation is suitable for 'whether' questions.

---

[10] Questions adapted from South East Sector Led Improvement Programme (2016) *Evaluation Framework for Early Help*: https://seslip.co.uk/download-file/440

[11] It is, however, possible to use a measure of distance travelled as part of an impact evaluation to assess a service's impact on an intermediate outcome. See Principle 4: Use high-quality measures.

Here are some questions that other kinds of evidence can answer:

- Do practitioners have the information and research they need to take informed decisions?
- What are the issues for families?
- Do all partners understand early help?
- Are we providing high-quality services?

Combining impact evaluations with other evidence

Many forms of evidence are useful in running services, and the best evidence bases involve multiple methods, qualitative and quantitative. A full evaluation of a service includes all of:

- **impact evaluation** to assess whether it works
- **process evaluation** to explore how and why it works
- **economic evaluation** to set out the costs and benefits.

What is a process evaluation?

Process evaluations do not answer questions about *whether* a service has a positive impact, but about *how* a programme's outcome or impact was achieved. They often explore issues such as how a service was implemented, the resources used, the practical problems encountered, and how such problems were resolved.

It is good practice to interweave process and impact evaluations, to help understand why a service is successful or unsuccessful: the process evaluation can help you understand the impact evaluation's findings on success or failure. Some local areas have been disappointed by a lack of conclusions from a process evaluation on whether a service is helping families or not – but for this, an impact evaluation is needed.

Process evaluations, sometimes known as implementation evaluations, can be qualitative or quantitative, but the best process evaluations combine both qualitative and quantitative methods. Sometimes process evaluations are carried out as feasibility studies of pilot services.

Process evaluations can draw on a range of data, including evidence on feasibility or how acceptable the service is to parents. Information gathered from staff can also be useful. For example, an impact evaluation might find that a service is not effective, and the process evaluation might find this was because staff failed to randomise participants to the intervention and control groups as intended.

The data used in process evaluations can be collected specially, or drawn from existing sources, such as staff surveys and management information on attendance and family demographics. Other useful information includes how much of a service families receive (dosage), drop-out rates and satisfaction scores.

Example of a process evaluation

The (then) Department for Communities and Local Government published a process evaluation in 2016 as part of the national Troubled Families Programme evaluation.[12] This involved case studies of 20 local authorities, chosen to ensure variety in the sample; interviews with 62 adults and children in 22 families in 10 of these local authorities, at the beginning and end of their involvement in the programme; and telephone interviews with 50 further local authorities. Among its findings was that appointing local Troubled Families Coordinators helped to achieve local buy-in for the programme.

[12] Department for Communities and Local Government (2016) *National evaluation of the Troubled Families Programme: process evaluation final report*: https://www.gov.uk/government/publications/national-evaluation-of-the-first-troubled-families-programme

Managers and commissioners need to combine evidence from evaluations with other research and monitoring data. This broader evidence base provides the answers to other essential questions: when does early help work, for whom, under what conditions, at what cost? An evaluation strategy can helpfully set out the types of evidence that will be gathered, to ensure answers are discovered to all the different questions services need to ask.

Why is early help difficult to evaluate?

Early help systems tend to be complex. They often involve more than one agency providing support to families with very diverse needs, and seek to influence multiple outcomes. Services often work with whole families rather than individual children or parents. And they provide many different services and interventions.

The complexity of many early help services is illustrated by a 2015 EIF survey completed by 26 local authorities. We asked respondents which services were involved in their multi-agency systems to provide early support to children, young people and families (see table 1). Sixteen service areas or agencies were cited by more than half of local areas.

TABLE 1

Service areas and agencies involved in multi-agency systems of early support

	Number of places (out of 26)	Percentage of places
Children's social care	25	96%
Troubled families	25	96%
Children's centres	24	92%
Youth offending	23	88%
Health visiting	22	85%
Police	20	77%
Schools	20	77%
Domestic violence and abuse service	19	73%
Other education services	19	73%
Child and adolescent mental health service	18	69%
Other mental health / psychological support services	18	69%
Community safety	17	65%
Early years providers (private, voluntary and independent)	17	65%
Maternity services	15	58%
Voluntary and community sector	14	54%
Housing	14	54%
Sexual health services	12	46%
NHS	11	42%
Adult services	11	42%

Source: EIF

This complexity means that impact evaluation is difficult. It is much easier to evaluate defined programmes or distinct interventions, such as new medicines, than whole services or complex community interventions. Indeed, some commentators believe the traditional medical model of linear development and evaluation of interventions is poorly suited to complex community interventions. Identifying causality in complex systems can be a challenge because many different factors can cause improvements – including changes in wider policy and practice, demography or economic circumstances – that are totally unrelated to the multi-agency system.

On the other hand, evaluating only the individual components of a system provides an incomplete analysis of the interactions in the whole.[13] Sometimes it can be hard to see the wood for the trees: through this guide, we want to help you to evaluate both.

The changing nature of early help provision makes it challenging to identify a fixed target to evaluate. Nearly half of local authorities responding to a 2018 survey by the Association of Directors of Children's Services stated that they had remodelled or changed their early help provision in the last two years.[14] It is difficult to separate changes caused by early help services from changes caused by other local changes and initiatives, including public service reforms, leadership changes, spending reductions and staffing changes. The early help offer itself is often also subject to changes in nature and extent, for example in response to funding cuts. This means that evaluations, while aiming to evaluate the impact of early help *provision*, may in fact be evaluating the impact of other changes.

All of this does not mean evaluation cannot be done, and done robustly.[15] It just means that more effort is needed to take account of the many factors that influence families.

[13] Department for International Development (2012) *Broadening the range of designs and methods for impact evaluations*: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf

[14] Association of Directors of Children's Services (2018) *Safeguarding Pressures phase 6*: https://adcs.org.uk/safeguarding/article/safeguarding-pressures-phase-6

[15] See for example: Medical Research Council (2006) *Developing and evaluating complex interventions*: https://mrc.ukri.org/documents/pdf/complex-interventions-guidance/

About this guide

The rest of this guide provides advice on how to apply the principles of good impact evaluation to complex local systems such as early help offers. The six principles are not arranged in order of importance or implementation; all will help to build a strong evidence base, as part of developing and continuously improving services.

- **Principle 1:** Know where you are starting from
- **Principle 2:** Prioritise outcomes to evaluate
- **Principle 3:** Embed evaluation in commissioning and practice
- **Principle 4:** Use high-quality measures
- **Principle 5:** Make comparisons
- **Principle 6:** Follow through.

These principles apply primarily to evidence generated locally by those involved in service delivery, but they are also relevant when thinking about commissioning an external evaluation. Most of the work needed to apply the principles will focus on making a service more evaluable – more ready for evaluation – rather than actually carrying out an evaluation. External evaluators cannot help here: only service commissioners and managers can make services more evaluable. However, local areas may need to look externally to find the expertise needed for some of the techniques that are discussed.

Each principle covers ground which should be achievable for all local areas, but also describes best practice. Following these principles will help a local area to improve its evidence base and begin the evaluation journey.[16]

[16] If there is a specific programme in your early help offer that you wish to evaluate, you should also consult our guide, *10 steps for evaluation success*: https://www.eif.org.uk/resource/10-steps-for-evaluation-success

Principle 1
Know where you are starting from

Summary

- You need a **good understanding of local families** if you are to evaluate the services they receive. The better your understanding, the better the evaluation will be.
- A good **needs assessment** should cover all local families as well as the subset who receive early help. It should go beyond describing needs at a point in time, to consider pathways of needs: how needs change over time for different families, strengths such as community links or strong relationships, and patterns of service use.
- As part of the **logic model** for your local offer, you need to understand the inputs, activities, outputs and outcomes involved, and the assumptions and external factors on which these depend.
- Carry out **system mapping**, involving partners and analysing data from partners, to set out the whole system, relationships and pathways through services, and how all local activities contribute to outcomes.

Why knowing where you are starting from is important

At a minimum, knowing where you are starting from means knowing what needs families in the local area have, which families are receiving services, and the range of services that are available. The better the understanding, the more able you are to take into account all the complex factors influencing family life. Having this ensures that an evaluation can ask meaningful questions of the right families about the real problems they face.

Assessing your area's needs

A good needs assessment starts with the baseline levels of need among the population in general, and among those who are coming to the attention of early help practitioners. It covers what these needs are, who has them, and how severe they are. For example, some areas use heat maps to simultaneously depict the seriousness and location of problems. You should consider the reach of your services: how much your services are used relative to local need, and patterns of service use, geographically and over time.

Data on local populations to inform needs assessment

A wide range of high-quality data is available for local areas to use as part of needs assessment, and to help in tracking progress towards common early help outcomes.

Units of analysis

Some data is available at the local authority level. Other data is presented at other levels (with other units of analysis), such as police forces or health bodies. And other data is only available at a regional level: for example, the proportion of households in problem debt, reported by the Office for National Statistics, based on the Wealth and Assets Survey.[17]

Calculating rates

It is always helpful to have basic numbers on hand about your local population's size and composition. Knowing the number of children, for example, allows you to calculate rates based on the absolute numbers (denominators), taken from other sources.

For example, the rate of emergency hospital admissions is the number of emergency admissions of children of a particular age, divided by the total number of children of that age. The number of children, young people, young women, young men, and births, can all be taken from the Office for National Statistics' mid-year population estimates. Similarly, the Annual Population Survey in Nomis gives the number of households containing children, and the Department for Education's school census gives the number of school pupils.

Understanding measures in context

For any data it is important to know:

- the organisation responsible for gathering it
- the quality of the data
- the exact question wording (where applicable)
- the frequency of updates
- the unit of analysis
- the coverage (for example, the age range covered)
- any limitations on the data (such as the fact that the data includes all adults, rather than only parents).

To ensure the information will be used, it is also helpful to keep track of the audience for the measure – the person or committee who will receive the report – as well as the target, last year's performance, and the person or group responsible for progress towards the target.

For example, rather than just tracking scores for a good level of development in the Early Years Foundation Stage Profile, the following information could be gathered and reported.

> **Name:** Early Years Foundation Stage Profile
>
> **Available at:** https://www.gov.uk/government/collections/statistics-early-years-foundation-stage-profile
>
> **Level:** Local authorities
>
> **Organisation:** Department for Education (although it may also be available at lower-level geographies, below local authority level, directly from local agencies)
>
> **Quality:** National statistics

[17] Office for National Statistics (2018) 'Household debt: wealth in Great Britain': https://www.ons.gov.uk/peoplepopulationandcommunity/personalandhouseholdfinances/incomeandwealth/datasets/householddebtwealthingreatbritain

Indicator: Percent of children achieving at least the expected level against all early learning goals (communication and language; physical development; personal, social and emotional development; literacy and mathematics), at the end of the academic year in which they turn 5.

Frequency: Annual

Limitations: Teacher assessment only. Coverage is high but not complete; for example, children are excluded who have not been assessed due to long periods of absence due to prolonged illness. Changes in measurement limit the comparability of years of data. Published data is not broken down below local authority level. The narrow range of scores limits the degree to which children can be distinguished from one another.

Value, England: 71.5% (2018), (2017: 70.7%)

Value, _____ (*your location***):** ___% (2018), (2017: ___%)

Target: ___% in 20___ (*year*)

Person or committee responsible: _____

Audience: _____

Much information is available about the populations of local areas. Everything from breastfeeding rates to GCSE results is freely available online, to download at local authority level. They are typically cross-sectional – that is, describing needs and outcomes at a point in time. Sources include Public Health England's Public Health Outcomes Framework, the Office for National Statistics' Nomis, and NHS Digital's Mental Health Bulletin.

Sources of data

The Public Health England profiles offer local data on the following topics:

- absence from school
- A&E attendances and hospital admissions
- alcohol abuse
- breastfeeding
- criminal justice system first-time entrants
- drug abuse
- employment and NEETs (Not in Education, Employment or Training)
- family homelessness
- free school meals
- immunisation
- poverty among children
- smoking among 15-year-olds and by women at time of birth
- under 16s and under 18s conceptions.

The Office for National Statistics publishes local data on:

- child sexual exploitation crimes and incidents (Crime in England and Wales)
- domestic abuse (Domestic abuse in England and Wales data tool)
- physical health and long-term conditions (Census)
- wellbeing (Personal wellbeing estimates)
- young carers (Census).

The Department for Education publishes local data on:

- childcare (take-up of funded places)
- children in need
- educational attainment (Early Years Foundation Stage Profile, key stages 1 and 2, GCSE)
- permanent and fixed-period exclusions
- special educational needs.

NHS Digital publishes local data on:

- mental health service use
- GP-diagnosed depression
- obesity, physical activity and diet
- self-harm and self-poisoning hospital admissions.

Published data is not available at local levels for the following topics, but may be obtainable from partner organisations in an early help system:

- antisocial behavior.

Published data is not available at local levels for the following topics, nor are partners likely to hold data on them, so early help offers with a focus on these topics may need to collect their own data:

- bullying
- child abuse and neglect
- family functioning
- gang membership
- interparental conflict
- parenting.

Thinking about needs in the round

It is important to understand both the nature of families' needs and how these evolve. Do some families get worse before they get better, with an initial dip before starting a journey of improvement? Do some families remain in need chronically, while others cycle in and out of need? How long does it take families to get back on their feet? Without the answers to these questions, gained from analysis of monthly data following families over time, it is not possible to identify the value added by services.

There is often a difference between incidence (new needs) and prevalence (total needs), so some families will newly experience problems that others have become used to over time. Knowing this helps you to design better-targeted services. It can also give confidence that the help provided to families is indeed early, not late, intervention.

For example, Lancashire County Council's Children and Families Wellbeing service received a referral from a school about a pupil whose attendance was poor. This problem had been going on for over a year, and the school were close to the point of taking action against the pupil's mother. The service considered this referral as having come too late, and would have preferred the school to raise the issue at an earlier point. It could work with schools to encourage earlier referrals, informed by data on the average duration of non-attendance among referrals from schools.

It is also important to understand patterns of service use. A good needs assessment covers the nature and coverage of existing public, private, and voluntary and community sector

provision. Do families step up and step down between service tiers, switching over time to more or less intense provision, and if so, which ones? What other services are being received? Some early help providers are unsure whether families are receiving services from other providers, so cannot isolate their impact on family outcomes.

For example, Better Start Bradford plans to bring together data from the 22 projects that make up the programme, to examine patterns of attendance, such as light or heavy users of services, and the characteristics of different groups of families with these patterns.

Broad-brush aggregate numbers are not enough. For example, the census will tell you the proportion of women of child-bearing age in your local area who are members of different ethnic groups. But some groups of women are more likely to have more children than other groups. So these proportions may not translate into the proportions of women in different ethnic groups among antenatal service users. Indeed, Better Start Bradford found that this was the case when its research team gained access to local maternity records.

Evaluators also need broader data on factors that may be masking the impact of early help services. For example, families living in isolated rural areas might not receive the same amount of early help as other families, and reducing social isolation might be harder to achieve. In this example, rurality influences the early help families receive, and also influences child and family outcomes.[18] Failure to consider this could lead to an incorrect conclusion on the effect of services for rural communities, with implications for service design to meet the needs of these communities.

Enrich your understanding by analysing differences between groups of local families

Bradford example

The Born in Bradford study is tracking the health and wellbeing of over 13,500 children born in the city between 2007 and 2011.[19] Data is being collected covering a broad range of health topics, such as weight, use of green spaces, and wellbeing. This enables a more in-depth understanding of Bradford's families than would otherwise be available. Born in Bradford found, for example, that south Asian families in the area had more fresh fruit and vegetables in their homes than homes of white families – but also more sugary drinks.[20] Findings like this have helped to develop a programme, HAPPY, to prevent the children of obese mums from developing childhood obesity. The programme has been developed in a way that will work with south Asian and white mums, for example by targeting the behaviours of the whole family, which is particularly important in those south Asian families where the wider family plays a role in shopping, cooking and feeding children.[21]

18 This is known as a confounding factor.

19 For more information see: www.borninbradford.nhs.uk

20 Born in Bradford (undated) *HAPPY: Promoting a Healthy Diet for Families*: https://borninbradford.nhs.uk/our-findings/different-findings-in-a-nutshell/happy-promoting-a-healthy-diet-for-families/

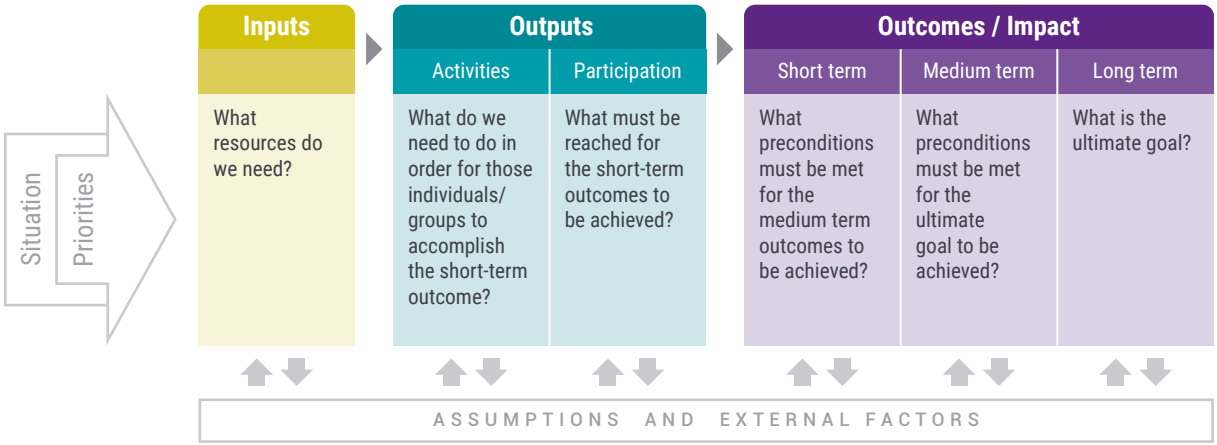
21 Taylor, N. J. et al (2013) Using intervention mapping to develop a culturally appropriate intervention to prevent childhood obesity: the HAPPY (Healthy and Active Parenting Programme for Early Years) study. *International Journal of Behavioral Nutrition and Physical Activity, 10*:142: https://ijbnpa.biomedcentral.com/articles/10.1186/1479-5868-10-142

Logic mapping: how and why you believe you are helping families

Needs assessment is the first ingredient in a theory of change and logic model. We recommend that programme developers set out a theory of change and a logic model as their first steps towards evaluation of their programmes.[22] A logic model is a statement of what a programme or service consists of and what it intends to achieve. It is a visual map of a programme's inputs, activities, outputs, and short and long-term outcomes (see figure 3).

FIGURE 3

A logic map of inputs, outputs, outcomes, assumptions and external factors



Source: EIF

Alongside this should be a statement of theory: how and why a service is expected to achieve its aims. A theory of change specifies why a service's intended outcomes are important, while a logic model specifies what the service will do to achieve them. A good theoretical understanding of how the service causes change will help you to identify and address any weak links in the causal chain.[23]

For an antenatal service, for example, your theory of change might assert that increasing pregnant women's knowledge of child-rearing, a short-term outcome, will lead to the longer-term outcome of improved child wellbeing.

It may seem that a theory of change and logic model are more about service design than service evaluation. But time and again evaluations have failed to verify the benefits of a service because they investigated the wrong outcomes, or because the core assumptions were wrong. Theories of change and logic models are essential foundations for evaluation.

In the broader context of early help systems, logic modelling and theories of change are more challenging. Nonetheless, you should develop a theory of change for the early help offer as a whole. This should bring together your definition of early help with your definition of success into a coherent statement of why it is that what you and your partners are doing is contributing to achieving your aims. It is important, especially with complex and wide-ranging early help offers, to map out the systems and processes in play, and what function you believe these serve.

[22] For more on theories of change and logic models, see our guide, *10 steps for evaluation success*: https://www.eif.org.uk/resource/10-steps-for-evaluation-success

[23] Medical Research Council (2006) *Developing and evaluating complex interventions*: https://mrc.ukri.org/documents/pdf/complex-interventions-guidance/

For example, consider the possible impacts of having a single front door as a feature of your offer. You might set out that the single front door means that families only have to tell their story once, making services more accessible, increasing satisfaction, minimising drop-out, and therefore maximising the proportion of families in need who receive help.

It is also worth considering what could go wrong. For example, you may not recruit enough staff or suitably qualified staff; time spent with families may be insufficient; or families may not take part, or put what they learn into practice. Considering these possibilities helps in process evaluations, which rely on theories or hypotheses about why a service works or does not work.

Often, the information needed for a logic model or theory of change is already available: it is held in the heads of staff and managers, and just hasn't been written down and agreed. Doing this work will help in linking the factors over which your local area does have control to the higher-level aims of early help. It may also help you to identify your area's lack of influence over some factors.

You cannot simply assert one-to-one links between services and improvements, but you should have a defensible basis for estimating the contribution you are making to families' lives. This is likely to be relatively modest. The evidence base for early intervention shows that the difference made by effective programmes is often quite small.[24] Nevertheless, though small, these are real improvements in families' lives.

Revise and refine your theory of change

Leicestershire example

Leicestershire County Council carried out a process evaluation of the early help work carried out with families between 2013 and 2017 by its case workers following an assessment. The motive for carrying out the evaluation was to understand family, partner and staff perspectives on strengths and areas for improvement, and how and why these vary for different types of families. The council's ambition is to expand evaluation out to other early help services, such as non-casework families who receive support in children's centres.

As part of the evaluation, Leicestershire developed a theory of change for its early help service. This was a key step, although the evaluation also included other methods, such as a staff survey and semi-structured interviews with 14 families. The theory of change helped to inform the design of the survey and interview questions, and in turn, the evaluation led to refinements in the theory of change.[25] As part of this, Leicestershire's evaluators carried out semi-structured interviews with parents to inform maps of their journeys, which they used to test the theory of change, to understand what was working well, and what could be improved. Staff were involved in workshops to inform the theory of change.

Map your system

A good understanding of where you are starting from includes a good shared understanding of the local system. The technique of system mapping can give you this. We see system mapping as aiming to identify the parts and relationships in a system that are currently available and important, those which are expected to change, how they are expected to

[24] For example, one of the programmes in EIF's Guidebook is a school-based programme called ParentCorps. It aims to increase parents' involvement in their children's education. The evaluation found that children who took part in the programme, relative to children who did not, had a higher average test score. The increase was 0.18 standard deviations, equivalent to saying 57% of ParentCorps children had a test score above that of the other children, compared to 50%, if the programme had no effect.

[25] Leicestershire County Council (2018) Early Help Evaluation Technical Report: http://www.lsr-online.org/reports,846738.html

change, and how to measure any changes. It can form part of an exercise to develop a theory of change and logic model, or be carried out separately.

The London Borough of Wandsworth has found system mapping an important tool in bringing together partners to discuss the components of early help that are currently available, the relationships between them, and the complexity of services. By using this to help develop a theory of change and logic model, they will be able to put in place the foundations for evaluating their plans to simplify their early help offer using whole-family approaches that focus on building resilience through the delivery of social and emotional learning.

System mapping can help you theorise about the relationships between services, and pathways for families. You should try to account for existing trends by understanding the wider factors that influence provision in your local area, compared to other areas. This will include the presence or absence of particular features (such as case work or group work) and the presence, absence or imbalances of partner agencies in the early help system, altogether and in particular districts or localities.

In practical terms, system mapping should involve a multi-agency discussion with the aim of developing a draft map as the output, for later refinement. The map should set out each component of the system – each service or interaction – and the relationships between them. It should not be carried out by a single person, but can be informed by a review of a small number of case files. System mapping can help to get partners to see themselves as part of a single system with a common purpose.

One challenge can be simultaneously representing different 'lenses' – such as settings, age groups and geographical areas – within a single system map. The more complex the local system (for example, the more vibrant the third sector), the more complex the mapping task will be. If it's helpful, multiple maps can be created to cover each lens of interest.

The benefits of system mapping include improving an evaluation, by helping you to explain the service and its context as part of an evaluation report. It can help you to reflect the reality of a system with components that interact and influence outcomes.[26] A good baseline understanding of relationships and pathways means that you will be well-placed to understand monitoring data. For example, it could help you to interpret a lack of referrals from a particular partner. This might represent either failure or success, if it turns out that the partner is carrying out effective early help on its own. A system mapping exercise involving all relevant stakeholders can help you to tease this out.

System mapping is a flexible method which can be adapted to local needs. It can be helpful, for example, to develop a map to add detail to each component, such as whether it is evidence-based or if it works with the whole family. Ultimately, the aim should be to combine the understanding gained from system mapping with data on the forms of provision and partners that families interact with. This may mean changes to IT systems to capture some information for the first time (for more, see Principle 4).

[26] Preskill, H. & Gopal, S. (2014) Evaluating complexity: propositions for improving practice: https://www.issuelab.org/resource/evaluating-complexity-propositions-for-improving-practice.html

Fill gaps and remove duplication between the services that make up the early help offer

Cheshire West and Chester example

The council's Early Help and Prevention service runs the evidence-based parenting programme Triple P for some of the parents its case workers help. The service is aware of other, separate provision: for example, the parenting programme 1-2-3 Magic for parents of primary-school-age children with behaviour problems (with or without ADHD). With the involvement of West Cheshire Child and Adolescent Mental Health Service, the 1-2-3 Magic programme is run by a variety of trained and accredited members of staff based at schools and the Countess of Chester Hospital. Other parenting services or programmes are being provided by others, such as housing associations, church groups, and the voluntary sector. Involving stakeholders in a system mapping exercise could address this gap in awareness. The council had been delivering 1-2-3 Magic too, but discontinued this provision due to the overlap with the mental health service's provision, thus removing duplication between services within the local early help offer.

Principle 2
Prioritise outcomes to evaluate

Summary

- **Set the scope** of your early help offer. You are only responsible for success or failure in achieving changes that are within the power of your early help system to bring about, and this is what you should evaluate.
- **Define success**, or 'what good looks like' if early help services succeed in helping families.
- **Be specific and realistic** in setting out a manageable number of desired outcomes.

Why prioritising outcomes to evaluate is important

Many early help systems are ambitious and comprehensive in their aims and are often designed to work with families to address *any* problem that they present with. If services are to be evaluated, however, it is important that commissioners decide the key aims of their early help offer or the most important aspects of the provision.

Having a scope and defined aims are prerequisites for evaluation: evaluators cannot evaluate everything that early help does. Without a statement of what is in and out of scope, it is impossible to know what to evaluate: where to collect data, from which people, asking what questions.

Set the scope of your local early help offer

The scope of early help varies in different areas. Some local areas define early help broadly, as an umbrella term encompassing virtually all local services provided for children and families by the local authority, other agencies, and the voluntary and community sector. Other areas have a single local authority programme badged as the early help offer.

This breadth should not be a problem for evaluators if the early help system is clearly defined, with agreed boundaries. It is for service managers, not evaluators, to set out the scope of a service. Partners should be involved in this process, as opinions may differ.

An early help evaluation may cover the whole offer, or individual services or projects; or build from these narrower areas of focus over time to cover a whole offer. It can be challenging to draw a boundary around an individual service or project when it is an integral part of wider provision. But for evaluation to be manageable, you may need to rule some of the complexity in the system out of scope, and acknowledge this restricted scope as a limitation in evaluation reports.

Define success: what does good look like?

As well as defining the scope of an early help system, it is important to determine its key aims. Many local areas have wide-ranging aims for their early help systems. Almost all the 28 local areas we surveyed in 2015 described the aims of their integrated system as including improved outcomes, positive user experience, and increased cost savings. Many local authorities are interested in reducing demand on children's services as an outcome of early help.

Be realistic

Your set of aims should be realistic, focusing on changes that are within the power of your early help system to bring about, and for which the partnership organisations are accountable. Early help systems cannot influence all aspects of families' lives.

Objectives for a local system should be feasible. You could target percentage reductions in particular needs, rather than aiming to reduce them to zero. For example, case workers can check that families are claiming all the benefits to which they are entitled, and so reduce child poverty in the area – but they cannot feasibly reduce to zero the number of children affected. While local authorities naturally want to be ambitious for children and families, overly ambitious targets make it more likely that an evaluation will find that they have not been reached. In particular, aims should extend to families who will be reached by the service, not all local families.

You should also consider what could go wrong, and what that could mean for family outcomes. Considering unintended negative outcomes helps in evaluations, because evaluators need to keep an open mind about possible positive and negative impacts, and may not have the knowledge needed to anticipate what the negatives might be.

Be specific

Aims should be concrete. Some areas have broad and loosely defined aims, such as 'service transformation' or that 'all children should thrive'. A more tangible way of expressing the latter ambition might be, for example, that children develop in line with expected milestones which are clearly set out across the different stages of child development.

It is difficult to evaluate high-level early help strategies containing aims and principles unless these are accompanied by concrete activities.[27] For each outcome area, you need at least one robust outcome measure (see Principle 4). For example, unless you have a valid and reliable scale to measure family functioning, then the outcome 'better family functioning' is too vague. A measure such as 'the absence of recorded incidents of domestic abuse for a six-month period' is more concrete.

When setting your aims, do not let the best be the enemy of the good: it is better to compromise and reach agreement with partners on a few key outcomes that are good enough, and to measure them well. You don't have to set out to be exhaustive. For example, Norfolk County Council's Early Childhood and Family Support service and Better Start Bradford both have a manageably small number of overarching desired outcomes, with three each. Bradford focuses on nutrition, socio-emotional development, and speech, language and learning.[28] Norfolk focuses on supporting children to achieve their developmental milestones; preventing more under-5s from experiencing neglect and emotional harm; and increasing social mobility.

---

27 For more information on how to develop a concrete 'blueprint' for a service see our guide, *10 steps for evaluation success*: https://www.eif.org.uk/resource/10-steps-for-evaluation-success

28 Better Start Bradford provides links to research on these three themes – see: https://betterstartbradford.org.uk/learning-resources/research-and-evidence/

Cluster themes in order to narrow down to a manageable number of priority areas

Lancashire example

Lancashire County Council's Children and Families Wellbeing service is ambitious, with a large number of desired outcomes and priority groups, and it recognised the need for greater focus if it was going to evaluate its impact. EIF facilitated a workshop to explore which outcomes the service should prioritise, and how these could be measured and evidenced. While senior managers suggested many priorities, these could be clustered into themes. For example, school readiness and early language both fell within an 'early years' theme. While this was just a first step, a single session was enough to reach a preliminary consensus on four key areas: the early years, risk-taking adolescents, family resilience, and parenting capacity.

The next steps for Lancashire will be to discuss and refine these priorities in consultation with stakeholders, including seeking consensus on the measures that should be used to track progress towards these outcomes, and setting expectations for movement on these measures. Many measures are available, but some are more suitable and higher quality than others (see Principle 4).

For example, the Early Years Foundation Stage requires a health and development review carried out when children are aged 2 to 2-and-a-half. This is a possible measure for tracking progress on Lancashire's early years priority. Then, the service could use this data to demonstrate that the children helped by the service are doing better on this measure than they were last year, or that they are doing better than comparable areas, or that the children they help are catching up with the general population of Lancashire toddlers. Any of these would involve comparison, which is a desirable element in any evaluation (see Principle 4).

Principle 3
Embed evaluation in commissioning and practice

Summary

- Rather than seeing evaluation as a one-off activity, try to **integrate evaluation** into the business-as-usual commissioning cycle.
- **Invest in evidence capacity** as part of continuing professional development: upskill staff on how to understand, interpret and gather data for evaluations.
- **Demonstrate leadership by using evidence**, and inculcate an evidence culture across the early help system.

Why embedding evaluation in commissioning and practice is important

Use of evidence is part of being a professional. Local areas should see evaluation and monitoring as part of continuous improvement, and prioritise evaluation as an important part of business-as-usual management and leadership. A high-performing local area would set expectations for what will be evaluated and the quality of evaluation that should be achieved.

When evaluations and services are commissioned in concert, it reduces the risk of evaluations running out of road. That is, it ensures that there is a service in place at the same time the evaluation is happening. Planning both services and evaluations at the same time, within the same timeframe, is best practice. It ensures that your evaluation will be relevant: that is, that it will be able to reach conclusions about what works, rather than what worked in the past, as part of previous provision.

Evaluation is part of commissioning

Ideally, evaluation should not be a one-off activity, but part of a cycle that repeats as circumstances change. It is part of the commissioning or policy cycle, and has a useful role to play at each stage.[29] Cycles – with clearly defined stages – make services more evaluable. Most policies and commissioning processes do not achieve this ideal, but there are ways to make it easier to integrate evaluation. Opportunities for evaluation can be exploited, such as when you are commissioning new services. And evaluation findings should inform later commissioning decisions about what to expand, reduce or adapt (as illustrated in figure 4).

[29] HM Treasury (2011) *The Magenta Book: Guidance for evaluation*: https://www.gov.uk/government/publications/the-magenta-book

FIGURE 4

Evaluation is part of the commissioning cycle

Source: EIF

At the same time, consider how to embed data collection into service delivery. For example, collection of 'before' measures can dovetail with existing assessment processes, and collection of 'after' measures can dovetail with existing processes designed to determine whether to continue or withdraw early help services after they have started.

Invest in evaluation

Prioritising evaluation means investing in evaluation. When evaluations are carried out internally rather than externally commissioned, the costs include staff and management time, training, and possibly the cost of software packages for data analysis. Although some software (such as R) is open-source, and a free data analysis add-in is available in Excel, many statistical software packages require expensive licences – and regardless of up-front costs, can require technical skills and expertise to use effectively.

Whether evaluations are conducted in-house or commissioned out, management time is needed to digest and communicate evaluation findings internally, and to track actions taken to implement recommendations. To do these things, managers need statistical literacy, and the skills to understand and interpret evaluation reports in particular, which take time and money to develop.

The resources needed for an evaluation vary: they should be proportionate to the scale, risk-level and profile of the service, and the existing evidence base. With respect to early help

systems, we know that evidence of impact is currently very limited, so there is a strong case for increasing and funding evaluation activity.

One rule of thumb, adopted for example by Better Start Bradford, is to assign 10% of a programme's budget to evaluation, or less in large-scale programmes. If commissioning out an external evaluation, local areas are unlikely to achieve good-quality results for under around £10,000 for a process evaluation or £30,000 for an impact evaluation. This is likely to be difficult to find, given current pressures on local authority spending, and central support for evaluation of early help systems is needed.

Build in plans for evaluation from the beginning

Hammersmith and Fulham example

The London Borough of Hammersmith and Fulham's Family Support service is a local authority-owned company which is providing early help, children's services and children's centre services. The service takes a trauma-based approach to providing support, aiming to support children to develop good executive function and resilience. It is developing a new programme, 'intensive family support', targeted at young people aged around 14 to 16 years old. It aims to improve school attendance as an intermediate outcome, with a long-term aim to improve young people's mental health and wellbeing. Although it is only at proposal stage, the service is planning how the new programme will be evaluated, alongside its work to specify the service model. Thinking ahead like this maximises the chance that the evaluation will provide useful information to help develop the programme.

Build a culture of evidence-based practice

Leadership is critical. Senior managers and early help leads should demonstrate leadership by championing evidence: routinely asking for the evidence to support any proposed change, and accepting the cost of evaluation as a necessary part of change. This entails properly resourcing continuing professional development (CPD), improving the availability and quality of management information, and gaining buy-in for funding for analytical teams. These steps together help to build a culture of evidence-based decision-making and practice in local services.

It is important to build evidence into the culture of a service: it is part of being a professional, and part of reflective practice. Training on evidence and evaluation can be built into initial staff training and CPD, to ensure consistency and high standards of professionalism across a service. It could also improve data quality: staff buy-in to data collection tends to increase if they understand the nature and importance of good-quality evidence and evaluation, which rely on good data.

Principle 4
Use high-quality measures

Summary

- **Exploit existing sources** of routinely collected data on outcome measures.
- Only gather data yourself where it is not already available. Where you have to gather data, **don't reinvent the wheel**: only use measures that have been carefully developed by experts.
- The measures you use should be: **valid and reliable**, relevant and responsive; reasonable in terms of length, simplicity and accessibility; allow tracking of progress or distance travelled; and broad enough to use with a range of families.
- **Consider your expectations** for the direction and amount of movement in the measures.
- **Improve data systems** to allow linkages between different systems and partners.
- **Consider introducing a shared unique identifier**, such as a unique pupil number or NHS number.

Why using high-quality measures is important

The term 'measures' can refer to anything from outcome indicators, like those published in the Public Health Outcomes Framework, to measurement tools, like the Strengths and Difficulties Questionnaire, which staff use to gather information from families.

Good-quality outcomes measures provide a narrative on success – for example, 'our early help offer has been shown to improve the school attendance of the families who benefit from it'. While output measures, such as the number of families reached, are also important, this part of the guide focuses on data on outcomes, which can be more challenging to gather. Outcomes data captures the overall aims of a service or intervention, and can tell you something about whether services are ultimately improving lives.

Where available, use existing high-quality measures

Evaluations should test whether there is a quantifiable improvement in at least one outcome that is a priority for the service, such as school attendance, child behaviour or levels of anxiety. This data can be secondary (already available) or primary, which means it is collected for the purposes of the evaluation itself. In practice, this is likely to be administrative data collected routinely and often required by government departments, such as Key Stage 1 outcomes.

Secondary sources of administrative data are useful for 'hard' outcomes, such as ceasing to offend or improving school attendance. For monitoring purposes, published aggregate data (headline numbers, abstracting from individual families) are enough. For evaluation

purposes, you may need to work with partners to gain access to the underlying, unpublished raw data so you can assess change at an individual or family level.

As a rule, we recommend exploiting existing sources where high-quality measures are available, and only gathering data yourself in cases where administrative data is an inadequate measure of the outcomes you are trying to influence.

Where no high-quality measures are available, identify valid and reliable measures

If your desired outcomes can't be captured using administrative data, then work with partners and other local areas at regional and national levels to encourage the use of high-quality, validated measures. Using the same measures as neighbouring areas enables you to make comparisons.

Evaluators should select measures that quantify what they are designed to measure, and that produce consistent results. You can tell whether this is the case for a particular measure, by checking whether it has been validated (see list of sources of validated measures on page 40).

Good-quality data which accurately represents the outcomes you want to achieve is essential. In short, nobody can have confidence in the findings and conclusions of studies that use measures which are not valid and reliable.

- **Invalid measures** do not measure what they claim to measure: they are not true measures of the concept you are interested in.[30] For example, measuring physical activity via a food diary is not valid.
- **Unreliable measures** do not accurately capture change. This is because they are unstable and vary randomly over time, rather than in line with the underlying change you want to measure. For example, a reliable measure of a person's level of empathy should be steady from test to test, unless something external – such as a programme – has changed their level.

There are some standard scales and templates which are widely used but do not provide useful evidence about impact. As well as not being valid and reliable, they involve subjective, quantified assessments by families of broad areas of life. For example, tools such as the Outcomes Star rely in part on practitioner perceptions (which are necessarily subjective) and do not represent a standardised measure – and therefore are not robust enough for evaluating impact.[31]

Validated measures have been carefully tested, which means they should be used in their entirety, without changing, adding or deleting any of the questions or answers. Establishing validity is a complex process, involving a series of statistical analyses to verify the measure. Using a 'pick and mix' approach provides no guarantee that the resulting, newly created questionnaire will still be valid.

Also, you should select measures that have been validated for use with a population resembling the sample in your study, particularly with respect to demographics and level of need.

Now, selecting measures that quantify what they are designed to measure, and that produce consistent results, is more easily said than done. Some measures, such as first-time offences, are straightforward, as they capture a relatively uncontroversial concept. But validated measures which seek to measure more contested concepts, such as child wellbeing or levels

[30] For more on validity and reliability, see our guide, *10 steps for evaluation success*: https://www.eif.org.uk/resource/10-steps-for-evaluation-success

[31] For more on the Outcomes Star, see the boxed text on page 39. See also: https://www.eif.org.uk/blog/choosing-the-right-measure-to-evidence-the-impact-of-early-intervention/

of parental conflict, can be lengthy, involve complex scoring, and require specialist skills – which means they can be difficult to administer within a busy service. And new measures can also take time to implement into practice, as they may require training for staff to use.

The team evaluating Better Start Bradford suggests the following steps to ease the process:

1. Review the fit between current measures and the intended outcomes.
2. Discuss a shortlist of validated, relevant, free or low-cost, and easy-to-use measures with practitioners.
3. Change databases and translate measures into other languages as needed.[32]
4. Train practitioners in the selected measures, ideally through someone regarded as an expert fellow practitioner, supported by a clear and comprehensive manual.
5. Pilot the new measures to test their feasibility and acceptability for families and practitioners, and the impact on time spent with clients and administration.
6. Implement the new measures, informed by the findings of the pilot.

Is it a measure of improvement?

It is important to ensure any selected measure captures information about *improvement* for a child or family. Some measures used in early help are ambiguous as to whether an increase would represent an improvement or a deterioration. For example, some local authorities track the number of young people referred to domestic abuse services more than once in a six-month period. An increase may represent failure, because the previous referral did not lead to the problem being solved. Alternatively, it may represent success, because it shows that awareness of the service is increasing and partners know how to make referrals.

Another example is case closure. If services have a fixed duration, then the number of case closures cannot be a measure of success. If, however, practitioners work with families for as long as necessary to resolve problems, then the time to case closure can be a measure of success.

Many local areas track entry into children's services as an indicator of the performance of their early help system. This has the benefit of being readily available, and allows the tracking of progress to some extent (through child-in-need status, to a child protection plan, to looked-after child status). It has the disadvantage of not being specific about the issue to be prevented or desired change. Often, it is through receiving early help services that more serious problems are uncovered, and so it would not be fair automatically to equate a higher number of escalations with an underperforming early help system. It is important to track relationships between parts of the system, while being careful in drawing conclusions.

Consider your expectations

You should also consider your expectations about the service's performance against your priority outcomes. How much difference would you expect to make to each measure…

- … if services perform at maximum effectiveness, under perfect conditions?
- … if services perform averagely, with some families dropping out or not engaging?
- … if services underperform?
- … in the absence of services?

[32] In our view, measures should be validated after they have been translated, to ensure that the translated tool continues to capture the intended concept. Many measures are already available in validated, translated versions for family members who do not speak English.

Don’t reinvent the wheel

A wide range of well-tested tools are available to capture many common risk and protective factors which are relevant to and upstream of common early help outcomes. A starting point should be exploring whether available tools could be used in measuring the impact of your system. It is not advisable to create a new measure or adapt an existing one. Creating a new tool runs the risk of wasted effort, as it is likely others will have previously developed and tested relevant tools, for use in assessing services with similar aims.

Where aggregate data is not available, you may have to collect your own data using validated measures. This is often the case for the outcomes many local areas are interested in measuring, such as bullying or the quality of parenting, which are not captured by administrative data. It can be helpful to use standard scales to measure, for instance, reduced aggression, wellbeing or better relationships. For example, a survey conducted as part of the national Troubled Families Programme evaluation asked main carers to complete two scales covering mental health and wellbeing: the short form of the Warwick Edinburgh Mental Well-Being Scale (see the boxed text below) and the Malaise Scale.

An example of a standardised scale

The short form of the Warwick Edinburgh Mental Well-Being Scale is a seven-item scale.

The next questions ask about your feelings and thoughts. Please say how often, if at all, you have felt each of the following in the past two weeks.

	1. None of the time	2. Rarely	3. Some of the time	4. Often	5. All of the time	*Don’t know*	*Don’t want to say*
I’ve been feeling optimistic about the future							
I’ve been feeling useful							
I’ve been feeling relaxed							
I’ve been dealing with problems well							
I’ve been thinking clearly							
I’ve been feeling close to other people							
I’ve been able to make up my own mind about things							

People score between seven and 35. A higher score indicates higher wellbeing.

In its evaluation of the Troubled Families Programme (2016), the (then) Department for Communities and Local Government found that mean scores for the main carer respondents questioned were very similar between those taking part in the programme and matched comparison groups (20.7 versus 21.1).[33]

[33] Department for Communities and Local Government (2016) *Evaluation of the Troubled Families Programme, Technical report: impact evaluation using survey data*: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/560503/Troubled_Families_Evaluation_Survey_Impact.pdf

Measure outcomes that you are accountable for influencing

It is not realistic to expect early help systems to impact on population-level measures, such as breastfeeding prevalence, if services only reach a small proportion of the local population. It may, however, be feasible to expect to see an improvement in breastfeeding rates among the families who are reached by early help services.

Aggregate indicators provide a useful overview of the population. But often, the measures are not granular enough to assess the impact of a service: for example, they might report economic inactivity for all adults, but not for adults who are parents, or parents who have been in contact with early help services. So it is worth exploring the potential for more suitable extracts and reporting with local partners. For example, you can consult NHS Digital's Mental Health Bulletin for the current proportion of your local authority's population who are in contact with mental health services.[34] But to report the proportion of parents who are in contact with the service, you might need to work with local mental health providers.

What are the problems with Outcomes Stars?

Outcomes Stars are widely used in early help systems as a method of evaluation. Outcomes Stars and similar frameworks can be a useful tool when working with families, but they are not a robust measure of impact.

This is because:

- They are not fully valid, reliable or norm referenced, which means they are not suitable for measuring change over time.
- They are generally used as part of a discussion between the family member and practitioner, and thus are not true self-report data.
- They are often administered on the same sheet of paper at both the 'before' and 'after' time points, meaning respondents may report an improvement in scores in order to please the questioner (which is known as social desirability bias).
- They are often vague in their wording, which does not lend itself to setting clear and reliable thresholds (for example, a cut-off score for whether a child is mentally unwell or not).

These tools could be used in a more concrete way. The outcome 'school attendance' could be measured using an outcomes star's 10-point scale, where school attendance (as a percentage) is captured in deciles for each school-age child in a family. This would be more meaningful than simply measuring a broad concept like 'school'.

Local areas can make best use of data already gathered using Outcomes Stars by analysing and interpreting the data as one piece of evidence about the early help system. It may be possible to explore areas of strengths and weakness in a service by considering how scores differ between different groups of parents. But this data should not be considered as evidence of impact.

[34] See: https://digital.nhs.uk/data-and-information/publications/statistical/mental-health-bulletin

Work with families to measure distance travelled towards outcomes

Many local authorities are interested in measuring the distance travelled by families towards outcomes, to know if services are helping people to make progress. There is a limited evidence base on what works in measuring distance travelled. Often, what local areas are interested in measuring is achievement of intermediate steps on the way to the ultimate outcome – such as regular school attendance as a step towards gaining good qualifications. But families may only travel a distance towards even an intermediate outcome, and not all the way, due to unrealistic assumptions about how long it takes for services to work. A well-informed theory of change and logic model (see Principle 1) can help you to determine how much time is required and plan data collection accordingly.

To some extent, families define success themselves. They work together with practitioners to solve issues in ways that are tailored, with personalised outcomes: a family may want to achieve fewer arguments, for example. Understanding the outcomes that families themselves are seeking to achieve is important. However, if these personalised outcomes do not align with the service's priority outcomes, they do not contribute to impact evaluation. Evaluators may be able to translate families' desired outcomes into the outcomes being evaluated: for example, from 'fewer arguments' to 'improved relationship quality'. Data on whether families have achieved their goals may need to be combined with more objective measures, focusing on the offer's priority outcomes.

Minimise bias

Many measures depend on self-reporting, and self-report data forms the basis of many of the evaluations of evidence-based programmes in our Guidebook. However, self-reporting can be inaccurate.[35] The quality of self-report data can be affected by issues such as poor literacy or lack of knowledge or motivation to provide accurate answers. Conversely, if measures are discussed and agreed with practitioners, as in the case of Outcomes Stars, there is a risk that scores are subjective to the point where they cannot be meaningfully compared. Externally-commissioned evaluations can avoid some of these issues by providing independent data collection, albeit at a cost.

To assess impact, it is essential to understand change over time by measuring the outcome before and after families receive services. This means asking the same questions, but not using the same sheet of paper each time. Some local areas use forms onto which families input data more than once. This kind of data collection is flawed, because previous answers can provide a prompt. As noted in the box above, family members may report an improvement, rather than the truth, in order to please the questioner.

Sources of validated measures

Free and paid-for tools are available, developed by academics, companies and charities. These cover topics as diverse as family beliefs, parental authority, homework problems, persistence, self-control, and schools, and children's relationships.[36]

---

[35] Institute for Employment Studies (2001) *Guide to Measuring Soft Outcomes and Distance Travelled*: http://webarchive.nationalarchives.gov.uk/20090902143103/http:/readingroom.lsc.gov.uk/lsc/SouthEast/distance1.pdf

[36] For further reading, see Family & Parenting Institute (2009) *Knowing what you do works: Measuring your own effectiveness with families, parents and children: a short guide*: https://tavistockrelationships.ac.uk/training-courses/practitioner-guides-resources/1134-knowing-what-you-do-works

Resources are available online to guide decisions, including databases of the thousands of measures that exist. There are a number of good sources of ideas.

- For parent–child interactions, EIF's report on what works to support parent-child interactions in the early years includes tables of validated measures, including for **children's attachment** (page 65), **children's behaviour** (p93), and **children's cognitive development** (p131).[37] One example is the Basic Achievement Skills Inventory.[38]
- For **child and adolescent non-academic and essential skills**, the Education Endowment Foundation provides the Spectrum database of measures with information on their technical quality and ease of implementation.[39] One example is the Strengths and Difficulties Questionnaire.[40]
- For **interparental conflict**, the EIF Commissioner Guide on Reducing Parental Conflict includes a list of validated measures used in evaluations of programmes with a component addressing the interparental relationship.[41] One example is the Parenting Scale.[42]
- For **child wellbeing**, **child mental health needs and family attributes**, the California Evidence-Based Clearinghouse publishes a list of reviewed measures, including assessments of how well validated they are.[43] One example is the Ages and Stages Questionnaire.[44]
- For **early years** services, the Education Endowment Foundation (EEF) provides a database of measures with information on their reliability and validation.[45] One example is the Child Behaviour Checklist.[46]

Consider the population norm for the measure

Some measures are **norm-referenced**, meaning the tool has been carefully tested on a carefully-selected group of test takers. This can provide local areas with a kind of counter-factual. That is, you can compare the population norm – a typical measure – to the value for the families in your evaluation, to see whether they are doing better or worse than average.

It is especially important to consider norms when looking at child development, as children grow and change between the start and end of an evaluation. Health visitors in Greater Manchester, for example, use the Ages and Stages Questionnaire (ASQ) multiple times as children grow through their early years. This is appropriate, because the ASQ comes in different versions for different ages, so change can be assessed by comparing children's relative standing at different time points.

---

37 Early Intervention Foundation (2016) *Foundations for life: What works to support parent-child interactions in the early years?* http://www.eif.org.uk/publication/foundations-for-life-what-works-to-support-parent-child-interaction-in-the-early-years/

38 See: https://www.familylearning.org/tests_basi.php

39 See: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluating-projects/measuring-essential-skills/spectrum-database/#closeSignup

40 See: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluating-projects/measuring-essential-skills/spectrum-database/strength-and-difficulties-questionnaire-parent-teacher-and-self-report-vers/

41 See: https://www.eif.org.uk/files/pdf/cg-rpc-3-3-examples-validated-measures.pdf

42 See: http://www.pti-sf.org/yahoo_site_admin/assets/docs/PS_English.242164902.pdf

43 The California Evidence-Based Clearinghouse for Child Welfare: Measurement Tools Highlighted on the CEBC: http://www.cebc4cw.org/assessment-tools/measurement-tools/

44 See: http://www.cebc4cw.org/assessment-tool/ages-and-stages-questionnaire/

45 See: https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluating-projects/early-years-measure-database/how-to-use-the-early-years-database/

46 http://www.aseba.org/forms/schoolagecbcl.pdf

Measure the unexpected

Good evaluators keep an open mind about impacts and watch out for unintended consequences – positive and negative. For example, an employability programme may have an unanticipated impact on parents' mental health. This is important in the early stages of developing a service and piloting its evaluation. It is also especially important given the large number of factors at play in a complex early help system.

How to pick a measurement tool

Which is the right measurement tool depends on the nature of the service, the families using it, and the desired outcomes. It should always be strategically appropriate, robust and practical, as described below.

Strategic fit

- **Relevance:** does it relate to one of the key challenges facing families in your local area?
- **Responsiveness:** is it within the power of your local area to influence?

Robustness

- **Validity:** has it been tested by others? Has it been found to capture the concept it claims to measure?
- **Reliability:** is it reliable, having been found to produce consistent results?
- **Norm referencing:** is it norm-referenced, meaning the tool has been tested on a carefully-selected group of test-takers, typically of the same age or school year and background?

Practicality

- **Length:** does it have a reasonable number of items?
- **Simplicity:** does it have a non-complex method of scoring?
- **Accessibility:** is training or professional accreditation needed for those administering the tool?
- **Trackable:** does it allow you to track progress towards outcomes (distance travelled)?
- **Breadth:** is it relevant to a broad range of families? Does the instrument cover multiple domains of childhood outcomes?[47]

Using the Strengths and Difficulties Questionnaire as part of an evidence-based parenting programme

Lancashire example

Lancashire County Council's Children and Families Wellbeing service runs a range of groups for local families. One of them is the parenting programme Triple P. As part of this programme, the service is gathering data from families using the Strengths and Difficulties Questionnaire, a validated measure which provides a brief behavioural screening questionnaire, which covers emotional symptoms, conduct problems, hyperactivity/inattention, peer relationship problems and prosocial behaviour. It has 25 items, and takes parents or young people five to 10 minutes to complete.

[47] Other questions to ask are included in Institute of Health Equity (2014) *Measuring what matters: technical report*, page 29: http://www.instituteofhealthequity.org/file-manager/measuring-what-matters-technical-report.pdf

One example of a question, for young people, is 'I get very angry and often lose my temper', with response options of 'Not true', 'Somewhat true' or 'Certainly true'. It is one of the most commonly used tools in early help systems, as it is freely available online as part of the Department of Health Framework for the Assessment of Children in Need and their Families (2000).[48] Thus, the service could use the tool in other settings beyond its Triple P programme, such as case work, to gain a broader perspective on the behavioural challenges facing the families it helps.

Measuring changes to ways of working

Many local areas have ambitions for their early help system that relate to transforming the system itself. As well as wanting to deliver outcomes for individual families, they want to improve the efficiency and effectiveness of services through, for example, improving relationships with partners, sharing data more effectively, or embedding a whole-family way of working. Unfortunately, evidence is not yet available on the impact of these changes to early help systems on family outcomes.

It can be difficult to know how to measure these changes. But many of the same considerations apply, as for measures relating to children and parents (see pages 21–23). Many suitable measures are available, including the number of referrals or assessments relating to each partner, attendance by partners at governance groups, and the amount of funding contributed to the partnership. You can also carry out a process evaluation (see page 15) to assess any changes to ways of working, using qualitative or quantitative methods, or both, starting with baseline measurements, and then repeating the same measures, after the changes to the ways of working.

Cheshire West and Chester council, for example, specifies in its Children and Young People's Plan that it will measure the proportion of referrals that come to it from partners and are accompanied by appropriate assessments tools, such as the Home Conditions Assessment tool.[49]

For most local areas, it is not realistic to go beyond these forms of measurement into impact evaluation territory, to discover the effect of the system changes on family outcomes.

Share unique identifiers across systems and partners

Evaluating a whole early help system might seem daunting on account of the disparate activities, programmes and forms of support included. But if it is genuinely a system, rather than an umbrella term for separate programmes, then there will be well-established links, and often a single front door. Evaluators can exploit this architecture to put in place data-gathering, if it does not already exist.

Beyond this, a more deeply integrated approach could be established, over time. More ambitious local areas could structure their information systems to allow analysis of families across databases and services. They could use standardised IDs, such as national insurance numbers, to allow matching and merging of data, in the same way that names and dates of birth are part of the data-matching work done locally as part of the Troubled Families Programme.

---

48 See: https://webarchive.nationalarchives.gov.uk/20130105133737/http:/www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4079384.pdf

49 See: http://westcheshirechildrenstrust.co.uk/wp-content/uploads/2017/05/West-Cheshire-Childrens-Trust-Review-2017.pdf

Working with practitioners, it is possible to standardise collection of key identifiers across services and partners. For example, this might be a unique pupil number or NHS number. The upfront investment needed to adapt records and modify databases would be compensated by the additional insight that can be gained from matching records together.

No early help system has yet achieved this level of integrated data, to our knowledge. We see two ways forward that would put a local area ahead of the pack: working with partners, and adopting evaluation-friendly systems when introducing new programmes.

Firstly, working with partners could improve linkages between data across an early help system. From the outset, the requirement to provide suitable data can be built into service level agreements or memoranda of understanding signed by early help partners.

Ethical considerations apply to quantitative research as well as qualitative research. Families should be told how their data is being used, and their personal data should be handled in accordance with the EU General Data Protection Regulation (GDPR). The data challenges that partners should seek to resolve in advance include:

- confidentiality, anonymity and data protection
- how data quality will be ensured
- data formats and systems
- procedures for data sharing.[50]

Secondly, new data systems can be designed to integrate with existing systems, to allow the tracking of family pathways between services. Be careful when adopting a new data system that forms part of a bought-in programme. Even if the programme is evidence-based, some local areas have found that the accompanying data system may not facilitate evaluation, because it does not capture the ID fields held in other local systems. This means that a manual matching exercise using names and dates of birth would be needed if the programme was to be included in an evaluation of the early help system.

Set an ambitious vision of shared information systems

Bradford example

Better Start Bradford is a lottery-funded 10-year programme (until 2025) led by a charity, Bradford Trident. It aims to improve early childhood development and bring about systems change.

In terms of data, Better Start Bradford's vision is of a single 'data spine' across health, social care, education and other local serves, but acknowledges this is likely to be years away. It recognises that information governance compliance across organisations, particularly in light of GDPR regulations, is challenging and complex. However, it also recognises the future benefits. For families, these would include less repetition of their stories, and, for partners, one-click referrals. Better Start Bradford is in the process of piloting a shared data system across different organisations, to improve the quality and consistency of data, and facilitate information sharing.

[50] Public Services Transformation Network (2014) Public Service Transformation: Introductory guide to evaluation: https://www.betterevaluation.org/sites/default/files/EvaluationGuideFinalv2.0.pdf

Build links between services into the IT platform

Cheshire West and Chester example

To support the work of its Early Help and Prevention Service, Cheshire West and Chester council uses a case management data system, Liquidlogic, with an early help module. Currently, health bodies and local police have access to the system, which is set up to allow tasks to be sent to the in-trays of colleagues in other organisations. Work is underway with partner agencies, such as schools, to widen access to the system.

Principle 5
Make comparisons

Summary

- **Collect data** from families both before they receive support from early help, and immediately afterwards.
- Make strenuous efforts to **follow up with as many families as possible**: even if they drop out of receiving services, they can still contribute data to the evaluation.
- **Test** whether the difference between the 'before' and 'after' data is statistically significant or likely to have been due to chance.
- Going further, **recruit or statistically construct an appropriate comparison group** who do not receive early help, ideally through randomisation.

Why making comparisons is important

Observing a positive change is a necessary (but not sufficient) requirement for finding that an early help service is effective. It is not sufficient because improved outcomes might have happened anyway: for example, a child's language delay may resolve itself naturally over time. 'Before' data (baseline or pre-service) and 'after' data (post-service) can only show an association between introducing a change and improving outcomes. At this stage, positive findings only indicate that the service is ready for more rigorous testing to estimate what would have happened in the absence of the service or programme.

What it means to make a comparison

Comparison is a desirable element in any evaluation. Most often, early help evaluation involves comparison over time, where families' outcomes before and after early help are compared. However, other types of comparison are common as well. You may compare your area against national or regional averages, statistical nearest neighbours, a specific local area believed to represent best practice, or a virtual, statistically created comparator. You can also compare performance against your area's past performance, or carry out internal comparisons of districts, wards, children's centres or schools, for example. Finally, different versions of a service can be compared, such as comparing high-intensity provision versus low-intensity provision, as in the Cheshire example on page 51. Each of these types of comparison is discussed in more detail in this chapter.

There is a place for all these types of comparison, as part of the evidence base for understanding an early help offer's performance. Whatever the type of comparison, the key principle is that the comparators should be genuinely comparable, meaning they are similar if not identical in relevant ways. For example, levels of deprivation should be broadly similar. But for a robust impact evaluation, two types of comparison are needed: comparisons over

time – before and after early help – and comparisons between two groups of families: those who do, and those who do not, receive early help.

Track change over time

To show there has been a change, people must be tracked over time. In doing so, there are some rules to keep in mind.

People should not be asked 'before' questions after they have received early help. They may not recall accurately, and there may be factors which distort their recollections. For example, asking parents to recall past emotional states, such as how confident they felt before they attended a parenting course, results in a non-robust counterfactual. Some baseline measures can be gathered retrospectively, such as employment or benefit receipt, using administrative records. These records were collected contemporaneously, and so they are free from recall bias. But most baseline data collection should take place before early help services begin, early in the assessment process. So, at a minimum, an evaluation should report on pre-service and post-service outcomes.[51]

Data should be collected from the same people at all time points. Recruiting a new sample each time does not provide confidence that change is real. This principle of linking data also applies to monitoring: you should know which groups of families have which attendance and satisfaction rates, for example. It is not good enough to draw on population-level data in evaluations, even if surveys are large and well-conducted. Without data from early help beneficiaries themselves, it is not possible to tell whether the early help offer caused improvements.

Comparing against previous performance

When comparing over time, it is important to look over a long-enough period to establish that two areas are similar, before tracking the impact of a change. For example, you might pilot a new service in half the districts in a local authority, and watch what happens to monthly data on outcomes in all districts over the course of the year, before, during and after the change.

Local authorities often make comparisons using this trend data, but care should be taken. A steady state may actually represent success, if you can demonstrate a robust counterfactual of rising costs or worsening outcomes in other areas. A substantial fall in cases could still represent under-performance, if other areas have seen even more substantial falls. Faced with a trend, you should ask yourself 'compared to what?' and consider what basis in evidence you have for your interpretation of the trend.

Benchmarking

Comparisons with other local areas can be enormously helpful. They can also mislead.

The right comparators vary from measure to measure: the best comparison for an early years outcome may not be the same as for outcomes relating to teenagers. It is important to gain an understanding of the context of the data you examine from other organisations, because data can be of poor quality. To understand your own data, and reasons for apparent strong or weak performance, you should have a good understanding of your local services and families (see Principle 1). Knowing what features of systems stay the same, and which features change, will inform benchmarking, allowing you to develop theories to explain observed trends. This knowledge need not be impressionistic, based on the views of service managers or practitioners, but could come from a process evaluation.

[51] For more information on pre/post studies, see our guide, *10 steps for evaluation success*: https://www.eif.org.uk/resource/10-steps-for-evaluation-success

It is also important not to credit changes in headline numbers entirely to your own efforts. The children's services caseload, for example, is affected by many factors, of which early help is only one. A theory of change and logic model (see Principle 2) will help you decide how much of an improvement it is reasonable to attribute to the early help system.

Managing drop-out of families: drop-out from services

Some families only attend some rather than all sessions of a programme. These families may be:

- **difficult to help**, because they are reluctant to engage with services
- **easy to help**, and drop out because they no longer feel the need for any help
- **average,** or drop out for neutral reasons, such as moving out of the area.

Do not pretend that drop-out from programmes has not happened and exclude these cases from analysis: evaluations should be of services as they are really delivered, not of a theoretical ideal world.

An 'intention-to-treat' analysis compares all those in the intervention and control groups, regardless of whether they stay in the programme or drop out. This kind of analysis produces more robust estimates of the impact for the entire cohort who entered the programme. For example, parents who attend all of an obesity reduction programme may be more motivated to lose weight than those who drop out or only attend a few sessions. In this case, removing those who receive a lower dose of the programme (who attend fewer sessions) may mean that your intervention group is made up of only the most motivated participants. This could falsely make the programme look effective by comparison with the control group, due only to differences in motivation between the groups rather than any impact of the programme itself.

Managing drop-out of families: drop-out from data collection

Whether they take part in services or not, all families should be followed up and strenuous attempts made to collect data from them. This is not easy – but if there is high drop-out, it is not possible to say whether the results reflect the true impact of a service or differences between groups that have been introduced by attrition. If drop-out from the study is too high (above 40%) then it cannot reach EIF level 2, which indicates at least 'preliminary evidence' of impact (see the annex for more on EIF's evidence standards). Some areas don't know much about what happens to families after early help ends, so cannot have peace of mind that problems are sustainably resolved.

Good-quality evaluations take steps to ensure people remain engaged at every data collection point. Such actions can include:

- clear communication of the benefits of taking part
- case management, such as assigning research team members to follow up with families
- maintaining detailed contact information, to boost the chance of tracking everyone down
- compensation, such as cash, vouchers or equivalent gifts
- reminding people, by letter, phone, email or text
- ensuring data collection is proportional and not burdensome.[52]

[52] Adapted from Brueton, V., Tierney, J., Stenning, S., Nazareth, I., Meredith, S., Harding, S., & Rait, G. (2011). Strategies to reduce attrition in randomised trials. *Trials, 12*(1), A128.

Consider collecting data at more than two time points

The characteristics of early help mean it can be useful to collect data while a family is receiving services, as well as before and after. An improvement at one time point is not enough, as there can be blips in data – trends must be sustained if you are to base claims of positive impact on them. Administrative data, such as monthly data on school attendance, can enhance data specially collected from families, and show that trends are sustained.

Bear in mind that it might take some time working with families for services to build up to full effectiveness; families may get worse before they get better. For example, Lancashire County Council's Children and Family Wellbeing service told us this was sometimes the case for the families they work with, in that it can take time for enough trust to be built up between a family and a case worker, for a family member to disclose a problem or how serious a problem is. In this way, services may intensify over time, rather than tailing off.

In practice this may mean that an immediate problem, such as a threat of eviction, is solved, before other important but less urgent problems, like indebtedness, are addressed. Uncovering these extra problems, or bringing them to salience, can mean extra work. This may go against the grain of many early help services, which have managing overall demand for statutory services as an objective.

Bear in mind that success may mean a problem staying the same rather than becoming worse, or a smaller increase in the proportion of families experiencing the problem. A number of the evidence-based programmes in the EIF Guidebook are based on evidence of this kind. For example, young people become more likely to be smokers as they move through their teens. So, for the programme ASSIST, success was marked by a slowdown in the worsening of smoking prevalence among young people in schools which received the programme, rather than a reduction in smoking prevalence.[53]

Test any differences you find statistically

It is important to analyse, not just report, the difference between 'before' and 'after' data. There is always a risk that the difference could be due to chance: statistical tests allow you to check this likelihood. Even if a difference is there, if it is not a statistically significant difference then it is ineligible for an EIF level 2 rating, which indicates at least 'preliminary evidence' of impact. See further reading in the annex for more explanation of the theory and practice of statistical tests.

Invest in upfront work to clean and link data before analysis begins

Greater Manchester Combined Authority example

There is a need for local evaluation of national programmes like the Troubled Families Programme. The national programme evaluation employs a robust quasi-experimental design, but it aggregates the effect of each local programme into overall effects for the whole of England. As local areas provide different kinds of programmes under the Troubled Families banner, it is important to break down the overall findings so that local areas can understand the impact of their specific services and whether they are improving outcomes for their population. The Greater Manchester Combined Authority (GMCA) has attempted to do this.

The authority's analysts undertook a lengthy exercise to link and clean data on families living in Greater Manchester who fulfilled the criteria for taking part in the national programme. This linking and cleaning was important, to remove duplicates and incorrect data, and so ensure a high-quality analysis. GMCA took this combined data and linked it to local administrative

[53] See: https://guidebook.eif.org.uk/programme/assist

data on child protection, school attendance and policing. They measured outcomes for each family 12 months before the start date of the programme, and 12 months after its end date. Adding together several years of data meant that GMCA benefited from a large sample size.

Initial findings were mixed, with some outcomes getting worse and some outcomes getting better. GMCA is using Excel software for its analysis. Excel has a data analysis add-in enabling the use of t-tests to check whether differences in means between two groups (the pre-programme group and the post-programme group) are statistically significant.[54]

To progress from this preliminary evidence towards evidence of efficacy (equivalent to EIF level 3), GMCA would need to identify a suitable counterfactual. In the GMCA context, the most promising avenue for identifying a comparison group is likely to be gathering more data from the 10 Greater Manchester local authorities on local families, including those with similar characteristics to troubled families but who do not technically meet the criteria for participation in the national programme.

Compare 'with and without': families who receive early help versus those who do not

Children grow in maturity, and with experience parents may gain in confidence and skill in parenting. So strong evaluation designs are needed to avoid identifying as 'impacts' benefits to families that would have happened anyway. A control group of families who do not receive services is needed for comparison with the 'intervention group' who do receive early help.

A comparison group is also important due to regression to the mean. Families may be referred to services at the moment when the problems are at their peak – and so if their needs are tracked over time from that point, they are very likely to show improvement. Outcomes may improve over time as an individual's run of bad luck comes to an end. Some of those entering enrolling into a drug treatment programme will have recently begun what turns out to be a temporary relapse. These kinds of 'natural improvements' could lead you to overestimate the effect of taking part in the programme.[55] Strong evaluation designs account for this by comparing improvements among those receiving services to those who do not.

Comparison groups exist naturally when there is a random element to service provision. For example, you might send parents to one of two versions of a service based on whether their birthdate is an even or odd number. A comparison group may receive 'business as usual' provision rather than the new service being piloted, or no provision at all.

There are ethical justifications for withholding early help from the control group, or offering a different form of provision, or a larger or smaller amount provision, compared to business as usual. After all, the alternative is providing *all* families with a service which may be ineffective. Communities should ultimately benefit from evaluation, which has improving services as one of its aims. This is especially true when it is used to test models that can be seen as having a reasonable chance of success, because they have good evidence from elsewhere and are based on a solid theory of change and logic model.

Concerns about withholding needed services can sometimes be addressed by piloting, or by using a 'waiting list' design, where families are randomly allocated to the waiting list or to a

[54] For more information on t-tests, see our guide, *10 steps for evaluation success*: https://www.eif.org.uk/resource/10-steps-for-evaluation-success

[55] Heckman, James J., and Jeffrey A. Smith. (1999) The pre-programme earnings dip and the determinants of participation in a social programme. Implications for simple programme evaluation strategies. *The Economic Journal 109.457*, 313–348.

service, and short-term outcomes are compared. This comes at the cost of not being able to track longer-term outcomes.

Evaluators sometimes exploit the fact that not everyone who is eligible ends up receiving services, and so avoid any active withholding of services from families. That is, they use waiting lists for oversubscribed services to create a 'without service' comparison group.

When using a comparison group you should prove that you are comparing apples with apples. Outcomes for the two groups should be found to be genuinely comparable over time, before this design can be considered robust. Without this, there is a risk that any change in outcomes might be due to differences in the demographics or other characteristics of the treatment group rather than the intervention. Their characteristics should also be broadly similar at both the pre and post time points.[56]

You should also minimise the risk of 'contamination', where the comparison group receive some of the early help service. Contamination is more likely in the context of early help, given the complexity of the system: it may be more difficult to know about or limit families' receipt of similar services through other pathways or agencies. Control for this in the analysis when it happens, if possible. One approach is to hold other factors constant by requiring other services to not work with families for the duration of a given service, as in the case of Multisystemic Therapy in Manchester City Council.[57]

Introducing a control group can seem daunting, but help is available (more information and further reading can be found in the annex). For example, the Better Start Bradford Innovation Hub recommends seeking expert academic support and advice when carrying out studies involving a comparison group. This can be achieved by establishing links with a local university. This may require give and take on both sides, but it has benefits: independent evaluations bring credibility.

Compare groups of families to each other, and compare outcomes before and after service changes

Cheshire West and Chester example

Cheshire West and Chester council's Early Help and Prevention Service helps families with a range of needs. It also helps adults affected by domestic abuse, whether or not they have children. It provides case work, and a range of programmes. The council recognises the risk that it cannot fully articulate the benefits of its services, and the risk that partners may not invest as a result, so has plans to improve its evidence base.

One of its interests is in why families come back to the service more than once. It has done some initial analysis of rates of return. Rather than looking only at headline numbers for all families in aggregate, the service has looked at rates of return for different groups of families:

- **Comparing two groups, difference in service received:** One of the ways the service has drilled down into the data is by comparing cases which have received a 'meaningful intervention' to those which have not. These could be seen as a higher dose and a lower dose. Return rates for those who have received a meaningful intervention, such as the evidence-based parenting programme Triple P, are much lower, although the service has not checked how similar or different the demographics of the two groups are, including their level of need, or whether the difference is statistically significant. It is working to improve its reporting, to allow it to extract data from its system on trends over time in this pattern.

[56] For more information on similarity between groups before early help (baseline equivalence) and after early help (differential attrition), see our guide, *Evaluating early intervention programmes: Six common pitfalls, and how to avoid them*: http://www.eif.org.uk/publication/evaluating-early-intervention-programmes-six-common-pitfalls-and-how-to-avoid-them/

[57] See: https://www.eif.org.uk/resource/multisystemic-therapy-manchester

- **Comparing two groups, difference between groups of families:** The service has found that re-referral rates are higher for domestic abuse cases than for other cases. It often sees a dynamic where a couple resumes their relationship and the abusive partner continues their behaviour. It also finds that perpetrators move on to new relationships and continue their behaviour, meaning the need for case work remains. Repeat rates can be up to 50%; however, this may indicate that victims feel confident enough to report the abuse.
- **Comparing over time, after a change:** The service has developed more close working arrangements with children's services than was the case in the past. Step-down from children's services to early help rather than broader partners is now considered, on completion of a single assessment or when risk has fallen after social work intervention. The council has noted a reduction in re-referrals of cases that have received a continuum of support. The council could improve its confidence that the change caused the fall by examining data on the trend over the period before and after the change, and considering whether any other service changes could explain the pattern. It should also analyse data on the characteristics of the families covered by the before and after data, to check whether they are comparable.

Principle 6
Follow through

Summary

- Evaluations are a tool for both accountability and for improvement. For accountability, you should **be open and transparent** by publishing your evaluation reports.
- **Be honest** and do not overclaim on the basis of your findings.
- **Provide enough detail** to allow someone elsewhere to replicate your findings.
- **Plan to develop and improve** the quality of your evidence and evaluations over time, to increase the robustness of your findings.
- **Act on your findings**, implementing the recommendations made in your evaluation reports.
- **Be patient**, waiting for numbers of cases to build up, and waiting for long-term outcomes to emerge.

Why following through on your evaluation is important

Evaluation provides an important evidence base for helping commissioners to decide how to improve their service offer and which services to continue, change, reduce or expand (see figure 5).

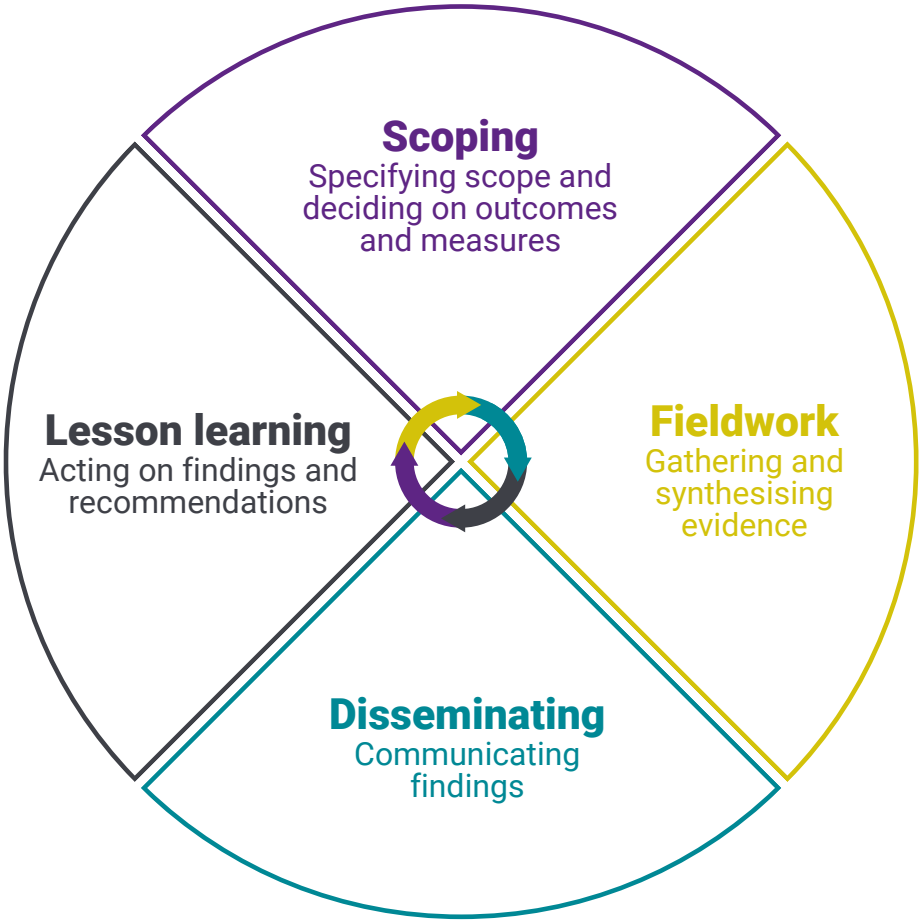
If your evaluation is high quality and its findings are favourable, you may want to consider expanding provision to help more families. If your findings are negative, this does not mean that a service will never work, but it does suggest the need to adapt and improve the service model, learning from the evaluation.

Whether your evaluation findings are positive or negative – or, more likely, mixed – there will be actionable findings about how different groups of families are affected, under what circumstances, and for how long. This should be taken forward through further research and evaluation, and through improvements to services. We see evaluation as a journey which improves in rigour at each stage, accompanied by refinements to the service model itself, informed by the developing findings.[58] This is especially important in the early help context, given that complex systems tend to develop and change over time.

[58] See our guide, *10 steps for evaluation success*: https://www.eif.org.uk/resource/10-steps-for-evaluation-success

FIGURE 5

The evaluation cycle



Source: EIF

Be honest and transparent

Good evaluations are transparent and acknowledge the limits of the analysis and reporting. This means clearly reporting all data and methods, and recognising any limitations, caveats and assumptions. If your evaluation design does not allow you to comment on causality, do not make causal claims. This means carefully wording your findings. For example, you might report on an *association* between receiving early help and people's positive outcomes.

In publications, make clear what procedures you have followed to assure yourself of the quality of your work. Good-practice quality assurance procedures include data validation, sensitivity tests and inviting others to scrutinise your findings, as in peer review. Peer review needn't entail publication of an article in an academic journal, but could involve, for example, recruiting a technical advisory panel. State which measures you have used, and details of any changes you have made to them. Some evaluation reports include the questionnaires used, reproduced in full in an appendix.

Evaluations are useful both for accountability and for learning. The accountability comes from placing in the public domain analysis of the impact of spending public money. It is also good practice to publish evaluation findings so that other local areas and future evaluators can learn from your experiences. The decision to publish or not should not rest on whether the results are favourable.

Plan in advance

Reporting needn't be a one-off event. To provide evidence to inform decision-making in the short term, it can be useful to plan interim reports as part of your evaluation timetable.

It is also good practice to publish evaluation plans in advance. Plans should cover what services will be evaluated, and when publication can be expected. Senior managers should be informed on progress throughout an evaluation. This should not just cover interim reports, but also progress updates, for example covering recruitment of participants and any changes to data systems that are needed.

During the planning stages you should think about the goals of your evaluation. What will be the uses for your evaluation findings? Who are the intended audiences? It is important to get stakeholders involved in these early stages, to decide together which outcomes are of interest and what evidence will be collected. If you include key stakeholders, such as commissioners and funders, in planning the evaluation, they are more likely to develop a sense of ownership and increase engagement in the process, helping to ensure that findings are subsequently acted on.

Manage expectations

Manage expectations of what research can teach us. Talk of 'impact evaluation' may be misleading, if the findings will not be able to draw conclusions about the impact of your service. You may need to frame the contribution to knowledge you can make, as part of a wider conversation to explain the nature of evidence and evaluation to stakeholders, colleagues and partners. This is particularly important given that, due to their complexity, we know that evaluations of systems of early help may not provide clear-cut answers. Similarly, commissioners and providers often expect clear results quickly. It is therefore useful to consider how you will report mixed and nuanced findings, and to be realistic about timeframes.

Horses for courses: who to report to

Given the wide-ranging nature of early help, it is particularly important to tailor versions of outputs to different audiences. It is also worth considering bringing partners together at the end of a project to share findings and agree next steps. One approach is to organise a validation workshop to discuss the findings and their interpretation with key stakeholders. Later on, it is good practice to publish a management response alongside the recommendations.

Audiences may include councillors, service managers, practitioners and funders. You may also wish to communicate your findings to the families who took part in your evaluation, and who expressed interest in receiving a summary. You should develop a plan for communication and dissemination to these audiences. This involves selecting an appropriate medium that is suitable for the content and the audience you are trying to reach.

What to report

Evaluation reports should describe features of the sample, evaluation design, methods, data analysis and findings, for the whole sample and for sub-groups of families.[59] Examples of the

[59] For a template for an evaluation report, see: https://project-oracle.com/resource-library/questionnaires

kinds of information to include can be seen in our 'Early intervention into action' case studies, covering East London, Fife, Manchester, North Tyneside, Sunderland and York.[60]

You should report all your findings. Even if the results are negative, it is still important to report and learn from them. Preventing others from making the same mistakes is a worthy legacy for an evaluation. A rigorous approach to reporting means that others can replicate the analysis, locally or elsewhere, even after the original evaluators have moved on.

The report should acknowledge the limitations of the findings, for instance, recognising that findings of a pilot study have low applicability to a wider context (low external validity) if the sample is not representative of the wider population.

Framing recommendations

You should work with stakeholders to draw up recommendations based on your evaluation findings. It is important to consider the caveats identified in the report. For instance, if you find that a pilot service has a positive effect but only for a subset of the population, you should not recommend extending the service to the entire population – at least not on the basis of this evaluation alone.

Recommendations should be SMART: specific, measurable, achievable, realistic and time-based. They should name the individuals or organisations responsible for implementing them. Processes should be in place to ensure the recommendations are in fact implemented, and you may have to follow up on these, or put in place regular reporting on progress. Recommendations should not only relate to service provision but also to evaluation itself: including suggestions for future evaluation and monitoring will help the smooth running of a cycle of learning and improvement.

Keep going

Evaluation has a natural cycle. Scoping gives way to fieldwork which, once evidence is analysed and written up, should lead to lessons to inform later evaluations. This is an ongoing process because services and circumstances are always changing, which creates the need to understand the effectiveness of new services or existing services in new circumstances. Like family life itself, evaluation is dynamic.

Include clear statements about caveats and limitations

Islington example

The London Borough of Islington's Early Help Family Support Services evaluation includes disclaimers to help the reader understand how they should interpret the findings. The report states that:

> *"This is NOT a statistically representative, scientific study on which hard conclusions about the overall effectiveness and impact of the different services can be reached. It is possible that that the evidence may have UNDER or OVER represented the benefits of support or the issues identified based on the sample."*[61]

As the evaluation was designed as a process evaluation, not an impact evaluation, this disclaimer is appropriate, to reduce the risk of readers misinterpreting the findings.

---

60 See: https://www.eif.org.uk/resource/early-intervention-into-action-innovation-and-evaluation

61 Interface Enterprises UK Limited (2015) *Evaluation of Islington's Early Help Family Support Services*: http://democracy.islington.gov.uk/documents/s4217/Early Help Executive Summary.pdf

Be patient

The successful delivery of early help depends on making a long-term investment, coordinated across all partners. This implies tracking long-term outcomes. The long-term benefits associated with effective early intervention include increased income and employment opportunities, reductions in crime and increased life expectancy. You have to be patient and plan ahead to capture longer-term outcomes. For example, Better Start Bradford plans to capture body mass index (BMI) at age 5, for the families recruited into its cohort study at antenatal clinics.

Collecting data on long-term outcomes gives a fuller picture: a service's short-term outcomes may reduce or 'fade out' over time, and there may be a delay between receiving services and effects materialising. While short-term outcomes are often considered important as steps towards the ultimate goal, it is still important to find out whether that ultimate goal has been achieved. Although it is costly and complex, we therefore recommend planning for long-term measurement, timed for when long-term impacts are expected to occur.

Do you have enough families to evaluate?

A further challenge for evaluators can be finding a big-enough sample, especially for small or specialist services. The right sample size depends on the design of the study and the size of the effect you are trying to identify: an estimate of how much good the service will do. This can be used to inform power calculations of how big the sample needs to be to answer the research question, using online calculators or statistical software. This requires realism about the likely impact of the service and likely drop-out.[62] To fulfil the EIF level 2 threshold, indicating at least 'preliminary evidence', a study must have a sample size of at least 20 (see the annex for more on EIF's evidence standards).

One solution is to extend evaluations over a longer time period, to gather data from a larger number of families. This will increase your sample size and make your findings more robust.[63] EIF considers a study to be robust (EIF level 3 or higher, indicating effectiveness or efficacy) if the final sample, after people have dropped out, contains at least 20 people in the intervention group and the control group. This is a minimum: a study of this size could only reliably detect effects that are very large.

Too few observations can mean your evaluation does not have enough statistical power to distinguish real improvements from improvements due to chance. The complexity of early help systems means more 'noise', so the 'signal' can get lost, unless you analyse data from a large number of families. This is particularly important to consider when evaluating light-touch early help services. The low dosage of light-touch services means a smaller impact, relative to other factors that influence families, and so a larger sample size may be needed – or indeed, data on all beneficiaries.

There is a rule of thumb. Imagine two variables in a table, with one as rows and the other as columns; for example, age group versus number of arrests. If you show the data like this, then the minimum number of observations (children or families) in a cell is 30. However, statistical software is often more lenient and will produce results for as few as five observations in a cell, and so these results should be viewed with caution.

---

62 Early Intervention Foundation (2018) *Evaluating early intervention programmes: Six common pitfalls, and how to avoid them*: http://www.eif.org.uk/publication/evaluating-early-intervention-programmes-six-common-pitfalls-and-how-to-avoid-them/

63 Medical Research Council (2008) *Developing and evaluating complex interventions*: https://mrc.ukri.org/documents/pdf/complex-interventions-guidance/

Follow up to find out whether initial effects are sustained

Essex example

In 2015, evaluators gained consent from 179 individuals to being tracked for evaluation purposes.[64] These people accessed Essex County Council's commissioned early intervention support for up to three months, at which point evaluators tracked their levels of engagement with their support. Later, at six- and 12-month intervals after support ended, evaluators followed up on individuals' access to specialist services, including for example referrals for domestic violence, drug and alcohol services, and adult mental health services. In addition to this tracking, the evaluation team interviewed 87 of the 179 people, to understand their views on the quality, impact and sustainability of the support they received. The tracking and interviews gave the council evidence on the sustainability of outcomes.

[64] Essex County Council, OPM group (2017) *Evaluation of the Essex County Council Family Innovation Fund: an early help programme*: http://www.outcomesstar.org.uk/wp-content/uploads/ESSEXFIF_EVALUATION_REPORT_FINAL.pdf

Annex

Questions for councillors to ask council officers about local evidence and evaluation

The following questions could be used by councillors to help them scrutinise the quality of evidence and evaluation available locally on an early help service, or could be used by officers as part of a self-assessment.

Evaluation of early help is at an early stage in most areas, so the answers to these questions are likely to be mostly negative. Their value is in improving understanding of the current situation and encouraging action to improve the state of evidence and evaluation locally.

Understanding local families

- Is there a good baseline understanding of the target population for early help?

Understanding the service

- Do we understand the whole local system, and what early help activities are going on across the local area?
- Is there a clear articulation of what our early help offer is trying to achieve and who it is aiming to support?
- Is there a clear justification for what services and practitioners are in our early help service and why?
- What is the evidence base for early help as it is practised here? Is the evidence applicable to this context?
- Do we have a set of SMART objectives and a set of measures for each objective? What is our theory of change for why our services cause the changes we want to see?

Understanding costs

- Do we know the full cost of our early help service and the number of people it works with per year? Can we track this over time?
- What is the budget for evaluation, in absolute terms and as a proportion of early help costs?

Understanding data, monitoring and management information

- Are suitable, high-quality data sources available, as part of business as usual?
- What data linking with partners in the early help system is carried out?
- Do we have data on how many parents, with what characteristics, are receiving early help? Where, and for how long? How satisfied they are with the service they receive? Have we used this data to analyse our reach?

Understanding impact

- What impact do our early help services have on the families we help?
- Is evaluation designed to enable assessment of the impact of early help on different groups of people? Are samples large enough to enable sub-group analysis?
- Is there a suitable counterfactual (a comparison or control group who do not receive early help) or could one be created?

Improving evaluation

- Are tools used in measurement valid, reliable and responsive; reasonable in terms of length, simplicity and accessibility; trackable, to allow tracking of progress or distance travelled; and broad enough for use with a range of families?
- Are we checking for unintended consequences?
- Are a range of qualitative and quantitative methods used?
- Is process evaluation carried out?
- Are evaluation results properly interpreted and contextualised to inform future service development and commissioning decisions?
- Is evaluation carried out to a high standard in terms of ethics, sample size and use of statistics?
- What are the limitations of the evaluations? What improvements are planned?

Embedding evaluation

- Is the offer/service designed so as to be evaluable?[65] Are changes piloted before being rolled out? Are there a manageable number of desired outcomes?
- What is the plan or strategy for future evaluation activity? Which services will be robustly evaluated, and which will just be subject to monitoring? What are the audiences and timescales for reporting?
- Is evaluation part of business as usual? Are analytical staff embedded in operational teams? Is data gathering for evaluation purposes embedded into assessment? Is training available to staff on evidence and evaluation? Do staff feel comfortable with evaluation?
- Are feedback loops in place to allow learning from evaluations? What changes have taken place based on previous findings?

[65] See Born in Bradford & Better Start Bradford (2017) *The Better Start Bradford Innovation Hub Framework for Monitoring & Evaluation*, evaluability checklist: https://borninbradford.nhs.uk/wp-content/uploads/The-Better-Start-Bradford-Innovation-Hub-Framework-for-Monitoring-and-Evaluation-V1.0.pdf

Further reading

- Early Intervention Foundation (2018) *Evaluating early intervention programmes: Six common pitfalls, and how to avoid them*: **http://www.eif.org.uk/publication/evaluating-early-intervention-programmes-six-common-pitfalls-and-how-to-avoid-them/**
- Early Intervention Foundation (2019) *10 steps for evaluation success*: **https://www.eif.org.uk/resource/10-steps-for-evaluation-success**
- Better Start Bradford Innovation Hub and Better Start Bradford (2017) Toolkit:
  - Service Design: An Operational Guide: **https://borninbradford.nhs.uk/wp-content/uploads/Service-Design-of-Early-Years-Inteventions-an-Operational-Guide-v1.0.pdf**
  - Implementation & Monitoring Toolkit: **https://borninbradford.nhs.uk/wp-content/uploads/An-implementation-Monitoring-Toolkit_V2.0.pdf**
  - Evaluation Framework: **https://borninbradford.nhs.uk/wp-content/uploads/The-Better-Start-Bradford-Innovation-Hub-Framework-for-Monitoring-and-Evaluation-V1.0.pdf**
- HM Treasury (2011) *The Magenta Book: Guidance for evaluation*: **https://www.gov.uk/government/publications/the-magenta-book**

///

Examples of evaluation designs and techniques[66]

Every evaluation is different, but broadly speaking evaluations can be grouped into a number of evaluation designs. This guide focuses on experimental designs, but other designs include theory-based approaches such as contribution analysis, 'case-based' approaches such as qualitative contribution analysis, and participatory approaches such as action research.[67]

Experimental designs include randomised control trials and quasi-experimental approaches.

- **Randomised control trials (RCTs)** are experiments where participants are randomly assigned to intervention and control groups so that they are equal on all known and unknown variables. Sometimes groups rather than individuals are assigned (for example, schools or regions).
- **Quasi-experimental designs (QEDs)** take advantage of the fact that individuals are exposed to the experimental and control conditions based on factors outside the control of evaluators, as when a policy design involves an element of random assignment (a natural experiment) or use statistics to deal with a lack of random assignment.

Specific quasi-experimental design techniques include:

- **Direct matching** (such as propensity score matching): members of the treatment group are matched with non-intervention group members to create a counterfactual that does not differ significantly from the intervention group on all known significant variables, such as demographic characteristics. Usually the two groups are separated by geography (for example, individuals who received a programme in a local authority may be matched to individuals in a separate authority where the programme is not available) or time (for

[66] Partly adapted from HM Treasury, Department of Energy & Climate Change, Department for Environment, Food & Rural Affairs (2012) *Quality in policy impact evaluation*: **https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/190984/Magenta_Book_quality_in_policy_impact_evaluation__QPIE_.pdf**

[67] Department for International Development (2012) *Broadening the range of designs and methods for impact evaluations*: **https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/67427/design-method-impact-eval.pdf**

example, individuals who received the programme may be matched to individuals in a previous cohort for whom it was not available).

- **Regression techniques:** a range of methodological approaches to understand the extent to which observed outcomes are associated with a variable, after controlling for relevant factors. Although commonly used in RCTs, regression techniques can also be applied to administrative, panel or longitudinal data. For example, in education we might use regression techniques on administrative data to understand the extent to which Ofsted ratings explain pupils' performance, after controlling for characteristics like socioeconomic status, gender, ethnicity and prior attainment.
- **Regression discontinuity design:** people are compared based on a cut-off in a pre-intervention measure. For example, comparing those who are 'only just eligible' with those who are 'not quite eligible' – as in the case where the effectiveness of free childcare is assessed by comparing those who were born just after the eligibility date with those born just before.
- **Difference-in-difference:** compares the change in the trend of the outcome between the intervention and comparison groups, before and after the intervention. This technique assumes that outcomes would have had the same trend in both groups had there been no intervention. For example, a policy designed to raise the school attainment of boys might be assessed by analysing how the gap between boys' and girls' attainment changes over time.
- **Instrumental variables:** the intervention and the comparison groups' memberships are allocated based on an external factor which influences the likelihood of being exposed to a programme, but which does not affect outcomes. For example, travel time to the programme venue might affect attendance but not parenting skill.

What good evaluation looks like for the Early Intervention Foundation

The Early Intervention Foundation assesses programmes against evidence standards (see figure 6) and publishes the results in the online EIF Guidebook.[68] This provides a way to judge the effectiveness of a programme, that is, whether evaluators have found the programme to have been effective in at least one rigorously conducted study. Our assessments cover both the impact and the cost of programmes. Some of the programmes in our Guidebook, such as variations on Triple P, form part of some local early help offers, and others have this potential.

Many early help services try to sustain their impact after services withdraw, so follow-up to assess longer-term outcomes is essential. The best evidence, scoring highest (level 4) recognises programmes with evidence of a long-term (one year or more) positive impact through multiple high-quality evaluations.[69] In short, a high-quality evaluation involves comparing a treatment group with a suitable counterfactual using intent-to-treat analysis.[70] This should be within a big enough sample for your findings to reach statistical significance; smaller sample sizes bring higher risks that findings are due to chance. As a rule, randomisation produces the most robust evidence.

[68] See: http://guidebook.eif.org.uk/

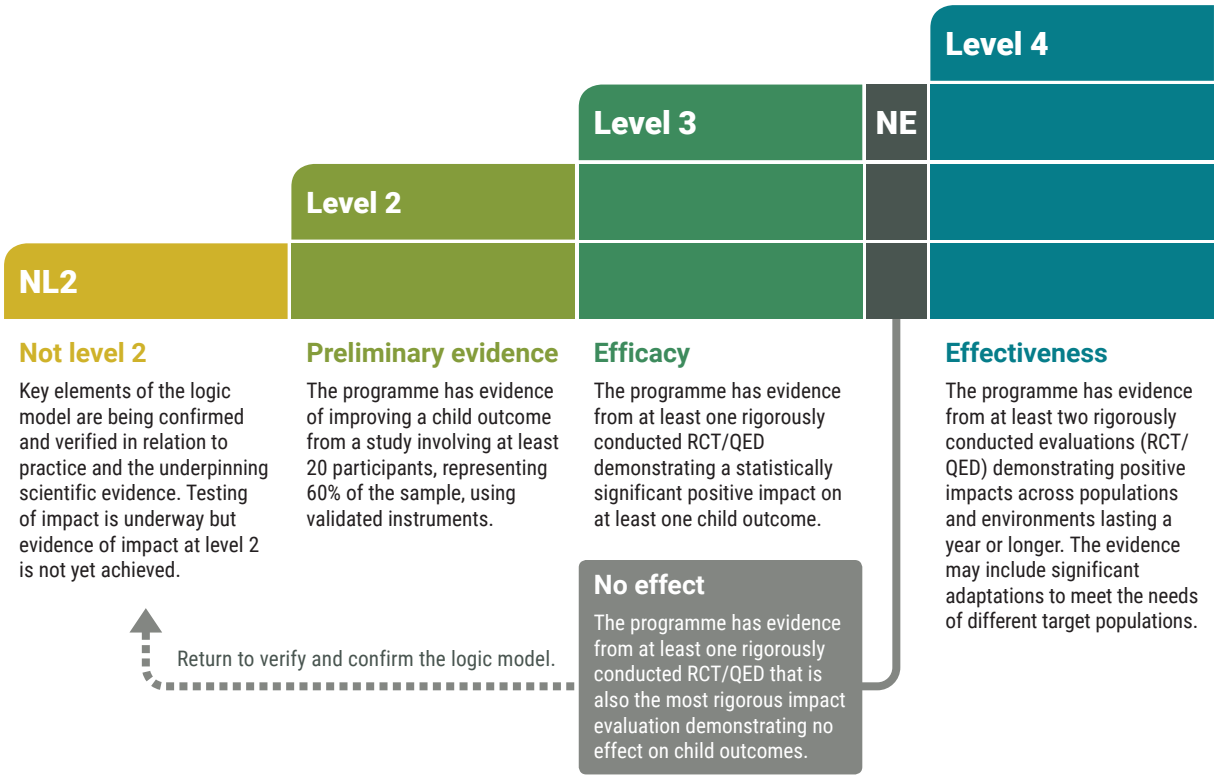
[69] For more about the definition and meaning of the EIF evidence standards, see: https://guidebook.eif.org.uk/eif-evidence-standards

[70] For an explanation of intent to treat, see Principle 3, Use high-quality measures.

Evaluations that use a pre/post design cannot receive an evidence rating higher than level 2 on the EIF scale. This is preliminary evidence only, rather than the level 3 and 4 evaluations needed to claim evidence of impact. Having both before and after data is an important step towards strong evaluation, and some local authorities are already gathering data in this way. But it is not in itself enough to make causal claims.

FIGURE 6

EIF evidence standards



Notes:
1. For an explanation of validated instruments, see Principle 3, Use high-quality measures.
2. For an explanation of logic models, see Principle 1, Know where you are starting from.