BALANCING ACT: A GUIDE TO PROPORTIONATE EVALUATION

Peter Harrison-Evans, Anne Kazimirski and Rosie McLeod

May 2016

CONTENTS

Introduction 3
Glossary 4
Evaluation: What is it good for 5
Focus and use of process and impact evaluation 6
Evidence standards: A guide for good evaluation? 8
Developing a proportionate approach to evaluation 10
Start with the theory underpinning your work 11
Look for relevant external evidence 12
Balance evidence needs with available resources 14
Tailor evaluation to the nature of your intervention 17
Choosing an evaluation design 19
Methodological considerations 19
Guidance table 20
Final considerations 23
References 24

Acknowledgements

NPC's approach to helping charities develop measurement and evaluation frameworks to assess their impact draws on the work and ideas of many organisations and individuals, the experience of our clients, and numerous evaluation and monitoring methods and models. It would be impossible to name all contributions individually. However, we would like to thank Lizzie Jordan and Peter Bailey at the Money Advice Service who commissioned and supported the paper. We would also like to thank our external reviewers David Pritchard and Bethia McNeil for their helpful comments and our colleagues at NPC for their valuable input.

INTRODUCTION

The last decade has seen a revolution in efforts to measure and understand the impact of social interventions and social sector organisations. As a result, far more evidence is now being produced on social sector activity. And perhaps more significantly, this has helped charities and funders think in a more impact-focused way, both demonstrating what they deliver and using data to help them improve.

However, the drive towards more and better evidence within the social sector can make some organisations feel like they are being pushed towards delivering high-cost, complex evaluations which are out of reach.[1] The prominence of evidence standards[2] for instance, has led to some organisations feeling that good evaluation practice means reaching the top of these hierarchies. On the other side of the coin, a rise in charities self-evaluating has led to some concerns over quality of evidence and the application of methods.[3]

To respond to these concerns, organisations need to identify ways of answering questions about how their services work and the difference they make with credible evidence, but without unnecessary spend. The key question, therefore, is what constitutes a proportionate and meaningful way to evaluate a programme, one that reflects an organisation's evidence needs, available resources and the operating context? It's a difficult, but crucial, balancing act to ensure an evaluation approach can be both sustainable and provide useful information for organisations to learn and improve, whilst also being accountable to stakeholders and funders.

About this paper

This paper aims to help you think through what proportionate and meaningful evaluation design looks like for your organisation. In doing so it draws on the literature on standards of evidence as one way of framing what good evaluation can be. It also argues that standards only help to answer some, not all of an organisation's considerations when determining their approach to evaluation. In the following sections we:

- Introduce the main types of evaluation and explore the role of evidence standards in assessing evaluation design.
- Outline the key considerations to inform your decisions about the type and level of evidence that is appropriate for you.
- Provide a guidance table as a reference document, linking evidence needs and research questions to methodological options that accommodate different situations and budgets.

This paper builds on NPC's previous work around impact measurement and evaluation. It draws on our four pillar approach[4] which guides charities through the key steps needed to develop a system for measuring the difference that make. However, this document takes a broader focus on evaluation (which includes impact measurement as well as other forms of evaluation—see table on page 6) and addresses wider questions that charities may ask when thinking about evaluation design. The guidance generally applies to both service delivery and to campaigns, though campaigning activity can be harder to evaluate[5]. Refer to the glossary at the start of the document for explanations of technical terms.

GLOSSARY

Attribution: Isolating and accurately estimating the contribution of an intervention to an outcome.

Baseline: A minimum or starting point in an intervention used for comparisons.

Causality: The relation between an event or events (cause or causes) and a second event or events (effect or effects), where it is understood the second is a consequence of the first.

Contribution analysis: A technique generating evidence on the role and influence of an intervention on an outcome, in the absence of experimental data.

Counterfactual: An estimate of what would have happened in the absence of the intervention or organisation.

Efficacy: The ability to produce a desired or intended result.

Evaluation: The use of information from monitoring and elsewhere to judge and understand the performance of an organisation or project.

Experimental approach: An evaluation that compares outcomes of recipients of an intervention to those of a control group who did not receive the intervention.

Fidelity: Is delivery faithful to the original model in terms of its core components, the duration of the intervention, and the quality of the service.

Impact: Usually the broad and/or long-term effects of a project's or organisation's activities, outputs and outcomes, after taking into consideration an estimate of what would likely have happened anyway (ie, the outcomes that can be reasonably attributed to a project or organisation).

Impact measurement: The set of practices through which an organisation establishes what difference its work makes. Can be used interchangeably with *impact evaluation*.

Meta-analysis: A study which looks at evidence across multiple studies

Monitoring: A systematic way to collect and record information to check progress and enable evaluation.

Outcomes: The changes, benefits, learning or other effects that result from what the project or organisation makes, offers or provides.

Outputs: Products, services or facilities that result from an organisation's or project's activities.

Quasi-experimental approach: A form of experimental evaluation where the control group is identified by matching the characteristics of people who are receiving the intervention using statistical techniques.

Randomised controlled trial (RCT): A study in which people are randomly allocated to one of several interventions.

Validity: The method which most accurately measures the issue you want it to.

Systematic review: A type of literature review that collects and critically analyses multiple studies or papers.

Qualitative: Research used to gain an understanding of underlying reasons, opinions, and motivation.

Qualitative Comparative Analysis: An analytic approach and set of research tools that combines detailed within-case analysis and cross-case comparisons.

Quantitative: Information or data based on quantities obtained using a quantifiable measurement process.

Pre and post: Research into beneficiaries and outcomes before and after the receipt of an intervention.

Triangulation: The use of multiple data source and types of data collection to test a particular research question.

EVALUATION: WHAT IS IT GOOD FOR

The many and varied approaches associated with evaluation can make it a daunting prospect for social sector organisations. However, at its core, evaluation is about helping organisations ask and answer some of the most basic and fundamental questions: *Are we making a difference? How can we improve? Are we reaching the right people? How does our programme lead to change?.*

It is sometimes assumed, however, that all evaluation boils down to one central question: *What works*?—certainly, there has been an increasing focus on this in policymaking.[6] 'What works' questions are important for understanding the effectiveness of an intervention and attributing impact to a particular activity or set of activities. But a narrow focus on 'what works' can limit the scope of evaluative research.

As well as answering questions on the attribution of impact, evaluation can be effectively deployed to explain the mechanisms of change (that is, why and how changes came about), and how they are influenced by the context of a programme or service. This provides different kinds of evidence which are needed to understand, improve, replicate or adapt a programme. These broader questions are well described by realist* academic Ray Pawson:

'What works, for whom, in what respects, to what extent, in what contexts, and how?'

Pawson and Tilley (1997) [7]

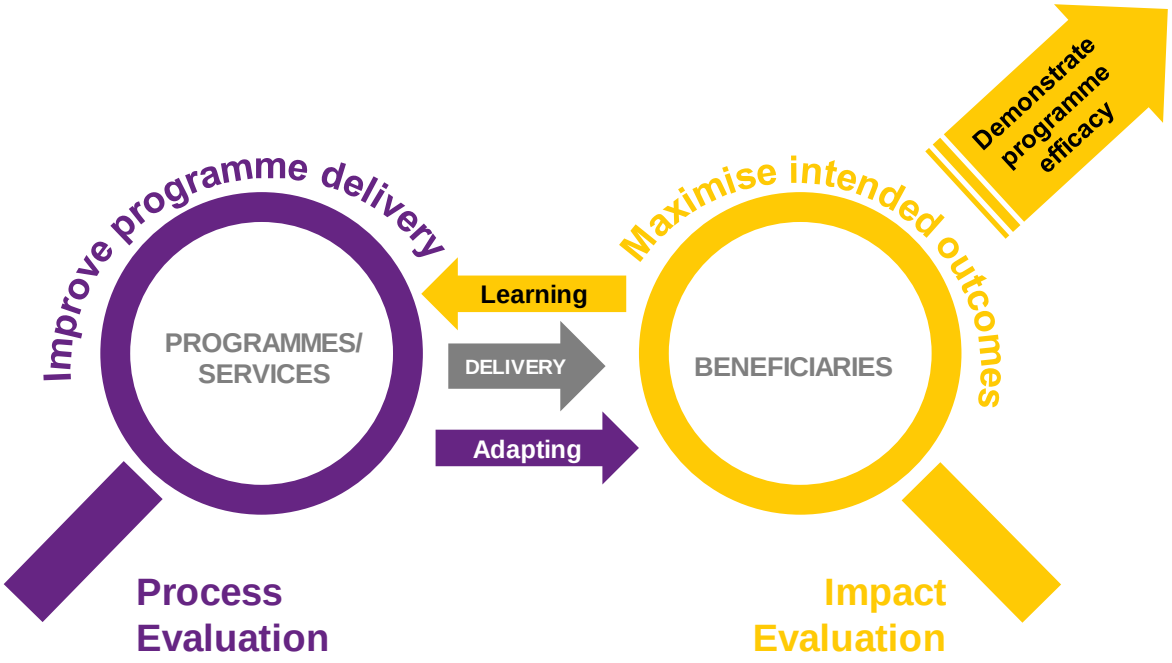
Fundamentally, evaluation is about providing a systematic approach to support learning: learning about what changes occur (descriptive); how and why change does or doesn't happen (explanatory and evaluative); and the causes of change or stasis (causal). Different approaches to evaluation can help provide different types of evidence and answer different kinds of research question. Often a mix of approaches is needed to answer relevant questions over the course of a programmes duration. The table on the next page summarises the main types.

* Realist/realistic approaches to evaluation are a form of theory-driven evaluation. Realist approaches stress the need for evaluation to understand mechanisms of change and context, information that experimental approaches generally fail to provide.

Evaluation type	When to use?	What is the focus?	Why is it useful?
Process	• Throughout programme delivery • Process evaluation can be *formative* ie, conducted on new programmes or services to inform delivery, or *summative* ie, conducted at the end of a programme or service codify the programme	• Developing a detailed understanding of programme operations • How successfully intended beneficiaries are being reached • How closely delivery is being implemented as planned (fidelity)	• Supports an understanding of how and in what contexts the programme is delivered best • Identifies ways to improve service design/programme delivery
Impact	• At defined intervals during programme delivery (can have formative and summative elements) • At the end of a programme	• Focus on assessing whether the intended changes have occurred for service users • Research that identifies the breadth and depth of change for service users • Attributing observed changes to programme activities • Can include economic analysis which measures the economic impact of a programme	• Provides evidence to demonstrate programme efficacy • Supports learning around how to maximise beneficiary outcomes • Economic analysis can provide information on the social and economic return on investment
Developmental	• In the early stages of developing a new social intervention/model • To support programme adaptation in fast changing and complex contexts (see p16 for discussion of complex settings)	• An iterative process assessing both programme delivery and indicators of impact on beneficiaries • Rapid and real time feedback that is linked to how the programme is delivered	• Provides methodological flexibility to support adaptation and learning • Supports social innovation in complex or uncertain contexts

Focus and use of process and impact evaluation

The diagram below illustrates the relationship between process evaluation and impact evaluation, and our reliance on both types of research to learn, adapt and improve a programme.



The information that evaluation provides can also be put to a range of different uses, including:

- **Designing services:** Trialling and refining different parts of a service to develop a model that seems to work.
- **Improving services:** Using evaluative learning to adapt and change (or in some cases stop) activities or services to maximise impact.
- **Demonstrating programme efficacy:** Answering questions about 'what works' and providing robust evidence on the success of a model at producing intended social outcomes.
- **Sharing ideas and building an evidence base:** Providing insight and knowledge about social problems and methods of tackling them, to support collective efforts around social change.
- **Maintaining accountability:** Evidencing to stakeholders how money, resources, and effort are supporting the aims of an organisation.

Any discussion of the virtues of evaluation must be accompanied by a cautionary note. Social sector organisations can feel pushed to over-claim and say that they have found definitive 'proof' that their intervention is effective. However, evidence can rarely provide a completely definitive answer to our questions. By its very nature evidence is 'partial, provisional and conditional'[8]—it may only be relevant to a particular context or time, however rigorously applied the methodology or well thought-out the design.

In addition, the situations we work in are constantly evolving, so our evidence needs change. This doesn't mean we should dismiss evidence or abandon our efforts, but it's important we think about the language we use when discussing our results. We should see evaluation and learning as continuous processes. This has been eloquently put by the team behind Realising Ambition, a programme of work to support the replication of evidence-based programmes:

'… too often we rely on evidence to support unequivocal claims of truth…even with the most robust evaluations, we cannot unequivocally conclude 'this works and this does not'. Rather, at best, we might be able to say that the evidence suggests that a particular intervention is effective (or ineffective) in improving one or more specified outcomes…'

Realising Ambition, Programme Insights: Issue 2[9]

What's the difference between research and evaluation?

Research is about generating knowledge about the world and how and why its works as it does. Evaluation is a form of research that is focused on understanding and improving social programmes. As a result evaluation specialists will share many similar methods, skill sets and ethical considerations as researchers in other fields.

Evaluation and wider research can be mutually reinforcing activities that help build an evidence base. Evaluation may for instance use research on levels of need within a local area to better understand a programme's service users. Similarly, evaluation can compare the changes observed in service users with research on population-level trends.

Evidence standards: A guide for good evaluation?

One way of thinking about impact evaluation design, which has risen in prominence in recent years, is 'standards' of evidence (eg, the Social Research Unit's standards of evidence, Nesta's adaptation of the SRU's standards, and the Maryland Scale of Scientific Rigour developed in the US).

These guides or hierarchies provide a helpful way to think through how the design of an evaluation, and the methods used, can influence the credibility of the evidence produced. However, to use standards correctly we need to understand the logic behind them. This includes understanding how they can help, but also knowing their limitations and what broader questions we need to ask to establish what proportionate evaluation design looks like for any programme.

Evidence standards generally place experimental approaches, such as Randomised Control Trials (or systematic reviews that summarise a number of RCTs) at the higher levels of their hierarchies. These approaches are favoured as they are seen as the most robust method of answering 'what works?' questions[10] that focus on causality and attributing impact to a particular intervention. Being able to compare results from 'treatment groups' (that receive the intervention) with 'control groups' (that don't) can provide strong evidence that the intervention was responsible for bringing about any observed changes.

This doesn't mean, however, that experimental approaches are appropriate in every context, that there aren't other ways to evidence causality[11], or even that causality should be the focus of your evaluation. A programme needs to be ready for this rigorous testing and the organisation needs to be able to undergo a time-limited, high resource evaluation. And despite the common logic behind many standards, all are focused on different priorities, sectors or stakeholders—reflecting the fact that there is no universal hierarchy of good evaluation design. To illustrate, the table below shows three different sets of standards:

Nesta Standards of Evidence[12]	EEF Security of Findings[13]	NPC's scale from *Untapped potential*[14]
Level 1: You can describe what you do and why it matters, logically, coherently and convincingly. **Level 2**: You capture data that shows positive change, but you cannot confirm you caused this. **Level 3**: You can demonstrate causality using a control or comparison group. **Level 4**: You have one or more independent replication evaluations that confirms these conclusions. **Level 5**: You have manuals, systems and procedures to ensure consistent replication and positive impact.	**1**: Comparison group with poor or no matching. **2**: Matched comparison, low statistical power (quasi-experiment). **3**: Well-matched comparison, medium statistical power (quasi-experiment). **4**: Fair and clear experimental design (RCT, RDD[†]) with weaker statistical power, greater attrition, weaker analysis & interpretation. **5**: Fair and clear experimental design (RCT) with greater statistical power and less attrition, stronger analysis & interpretation.	**Low***:* *Descriptive*—there is a clear explanation of how and why activities should lead to desired outcomes. **Medium***:* *Single method*—captures data using a single method that shows a change. **High**: *Mixed method*—captures data using mixed methods to triangulate a change. *Comparative*—assesses causation using a control or comparison group. *RCT*—assesses causation through the random allocation of the control and test subjects.

These standards share some similarities, with higher complexity comparative approaches at the higher ends of the hierarchies. However, they also exhibit some key differences:

- The Nesta standards aim to draw together different approaches to evaluation (experimental and theoretical) while also thinking about what evidence is useful to help organisations and investors assess the potential for scale.

[†] Regression Discontinuity Design: A type of a quasi-experimental research design where participants are assigned to the treatment and control groups on the basis of whether they meet a certain threshold. Source: https://educationendowmentfoundation.org.uk/evaluation/the-eefs-approach-to-evaluation/evaluation-glossary/

- The EEF scale is much more focused on experimental approaches, and differentiates between them in greater detail. This reflects the EEF's focused role in developing the scientific evidence base around education interventions.
- The NPC scale reflects its aim to assess charity evaluations (in the health sector), recognising that the high-cost research designs at the top of other hierarchies are often out-of-reach of individual voluntary sector organisations. As a result the scale looks to differentiate to a greater degree between approaches that would have sat at levels 2 and 3 in the Nesta standards.

Therefore evidence standards can help us think about what kinds of evidence different research designs can produce, and how different stakeholders will view the credibility of that evidence in terms of demonstrating impact. However, given their focus on particular evaluative questions ('what works') they are not designed to provide a comprehensive guide for organisations when thinking about what evaluation approach to take. In the following sections, we look to address those broader questions which will help you to understand the right approach for your programme or organisation.

Understanding quality in evaluation

How we can determine the quality of a study is not a question that is often covered by evidence hierarchies, but is of real importance when interpreting evidence and assessing its value. This goes beyond the type of evaluative approach taken and includes:

- Methodological quality—how well executed was the research?
- Appropriateness of methods to the aims of the study—does the method match the research questions and purpose of the study?
- Quality in reporting—is the research presented in a way that can be appraised and used by others?
- Relevance to policy and practice—does the research address important policy and practice questions in a way that is both useful and useable?[15]

With these criteria in mind, a poorly designed RCT could produce less credible (and useful) results than a rigorously applied before and after statistical analysis, for instance.[16]

Quality is, therefore, about adhering to good social research practice and understanding the context in which you are working. For example, Bond's evidence principles[17] focus on quality in the context of international development, providing a checklist for organisations to think through when reviewing or designing evaluations:

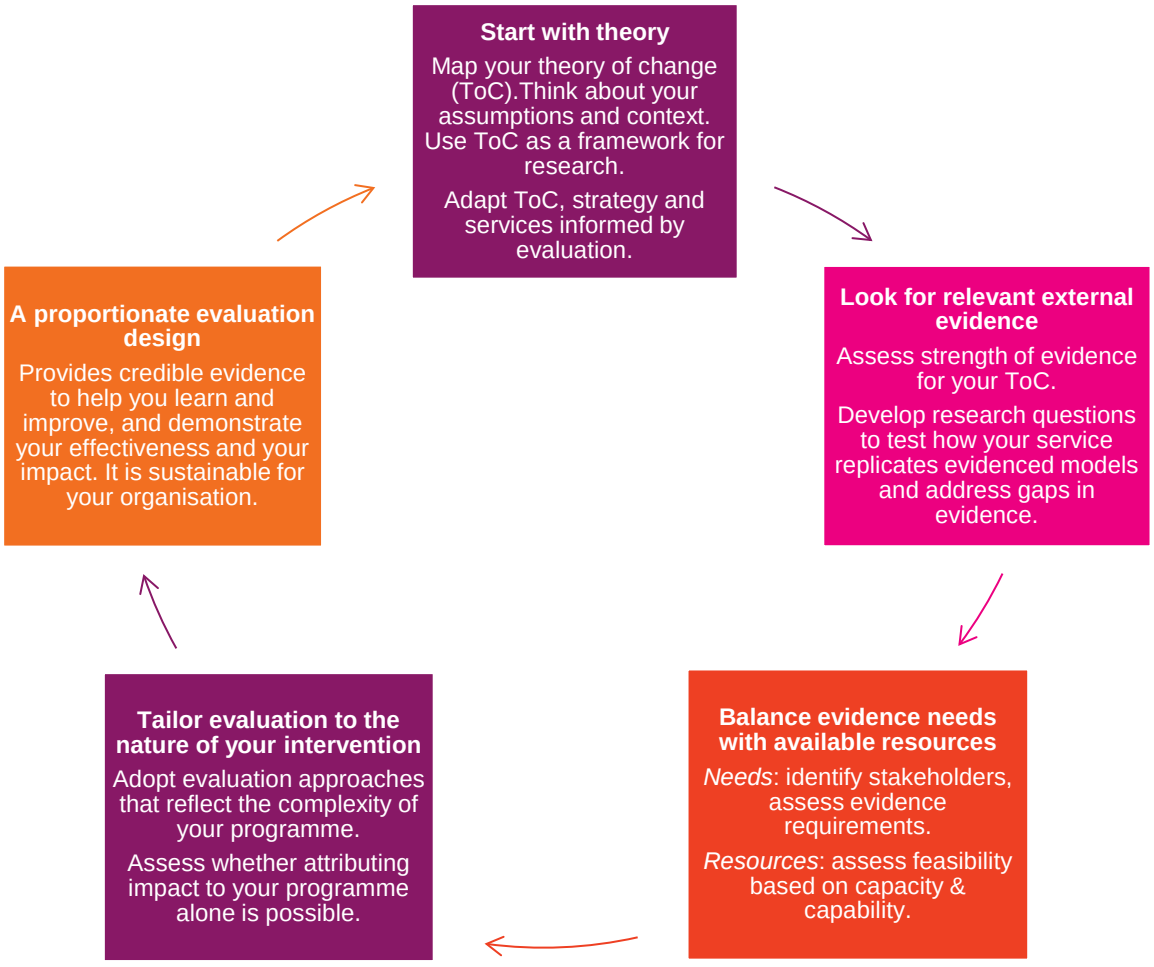
1. Voice & inclusion: The perspectives of people living in poverty, including the most marginalised, are included in the evidence, and a clear picture is provided of who is affected and how.
2. Appropriateness: The evidence is generated through methods that are justifiable given the nature of the purpose of the assessment.
3. Triangulation: Conclusions about the intervention's effects use a mix of methods, data sources and perspectives.
4. Contribution: The evidence explores how change happens and the contribution of the intervention and factors outside the intervention in explaining change.
5. Transparency: The evidence discloses the details of the data sources and methods used, the results achieved, and any limitations in the data or conclusions.

DEVELOPING A PROPORTIONATE APPROACH TO EVALUATION

A good research design is always built around the questions that are most important for the organisation to answer. While it's easy for organisations to be 'methods led'[18]—selecting data collection methods first (eg, a survey or interviews) and then basing evaluative work around this—this reverse engineering is a missed opportunity to develop an approach which provides the most useful kinds of evidence for your programme. What's more, taking a thorough approach to developing an evaluation design can prove cost efficient—targeting the most useful questions and methods and dispensing with data collection that doesn't serve a clear purpose.

The following section aims to help organisations think through what kinds of research questions will be most important for them, and the approaches that will help them to best answer these questions. It focuses on the criteria to consider—suggesting you start with the theory behind your approach, assess the existing evidence, balance your evidence needs with available resources, and tailor your evaluation to the complexity of your programme—to develop a proportionate evaluation design. These considerations are applied in the **guidance table** in the next section, which suggest the types of methods that may constitute a proportionate evaluation design, based on the external evidence, needs and resources, and the nature of the intervention.

Criteria for developing a proportionate evaluation approach



Collaboration and shared measurement

You do not need to develop an evidence base all by yourself. Collaboration between organisations working towards similar outcomes makes it easier for organisations to learn from each other and save on costs of developing tools. Pooling data means we can build a larger evidence base more quickly, which supports service improvement across the sector.[19]

The Financial Capability Outcome Frameworks[20] developed by NPC and MAS is an example of this collaborative work around outcomes. These frameworks—for adults, children and young people, and teachers—set out a shared language around financial capability outcomes, developed through a review of evidence and an extensive consultation with a large number of sector experts and practitioners. Using these frameworks can help organisations to articulate the ways in which their activities lead to improvements in the financial capability, behaviour and well-being of their clients, as well as to robustly and consistently measure the scale of the impact they achieve.

Start with the theory underpinning your work

On the guidance table: all methods recommended are underpinned by a theory of change.

- **Start by thinking about the theory behind your programme or service.**

This may seem a rather academic concept, but all interventions are based on theories, however subtly they are expressed.[21] These are the ideas or assumptions that underpin how you think your service works and why you think it achieves positive outcomes for your beneficiaries. For example, a mental health service provides beneficiaries with meaningful activities and non-judgemental company because they think this increases self-esteem and supports steps to employment.

- **Develop a theory of change**.

This is useful way to take a more systematic view of the theory that lies behind your programme. A theory of change is a description or map of how your activities link to your final goals through a series of intermediate steps.[22] This can help you to understand all of the important outcomes of your work, enable you to map how these outcomes are connected, and provide a way to track progress towards your final goal.

- **Use the theory of change as a framework to develop evaluation questions.**

At NPC we use theory of change as the basis for the majority of our evaluative work, as it helps define the focus of your evaluation and prevents you from attempting to measure everything.[24] This approach supports a gradual, cumulative generation of knowledge and the theory of change should be continually tested and revised over time, drawing on relevant external evidence, as well as through conducting primary research.

Developing a shared theory of change: NHS Charities

In situations where a number of organisations are working towards similar outcomes or are trying to tackle the same problem but from different angles, it may be possible to develop a theory of change at the sector level.

A group of NHS charities have recently taken this step, developing a theory of change which articulates a shared vision of the value of NHS charities to the health system. This has provided a framework to support collaboration around common outcomes as well as providing the basis for a shared system to demonstrate their effectiveness.[23]

Look for relevant external evidence

On the guidance table: column 1 shows the range in level of external evidence to match your situation to.

- **Assess how far your theory and assumptions are supported or contradicted by existing evidence.**

This will help you to determine the priorities for your evaluation and shape your research questions. If the approach is well evidenced there is no need to reinvent the wheel—you can focus on using evaluation to help you to refine your service and support improvements in service user outcomes. A strong external evidence base may also mean that stakeholders are less demanding about the amount and standard of evidence required in demonstrating your effectiveness.

- **Reviewing relevant evidence is a key part of the learning process.**

It can help you think about how others have delivered similar services and encountered similar challenges elsewhere. This in itself can provide valuable information in developing or adapting your services, even before you establish an approach to evaluation.

Drawing on existing evidence

- **Map existing evidence on to your theory of change.**

While it may not map neatly on to your theory of change in its entirety, it is likely that there will be some academic or 'grey literature'[‡] of relevance. The more the weight of external evidence can reasonably underpin your theory of change, the more confident you can be that your programme can achieve similar outcomes.

- **But make sure you don't cherry-pick.**

When looking for external evidence it can be tempting to select only the sources which support your model. However, this will give a distorted picture of the external evidence base and the extent to which your programme aligns with it. The most robust way to prevent cherry-picking is to conduct a systematic literature review[§], however, this is a lengthy and costly process, out of reach of most social sector organisations. If this isn't an option it is still important to look widely for potentially relevant evidence—including drawing on systematic reviews that have already been conducted—and present a balanced picture of what you find.

- **In reviewing external evidence, attention to context is key.**

Knowing that a model has delivered successful results in other settings cannot guarantee that it will do the same in yours. To understand how changing context could influence the strength of supporting external evidence, close attention needs to be paid to enabling factors within your theory of change—those elements which need to be in place for you to achieve outcomes (eg, the policy context, levels of need, delivery partners, and infrastructure).

- **Use the weight of existing evidence to begin to define the focus for your evaluation activity.**

It is difficult to judge how much additional research you need to conduct to be confident that your service should replicate others' results. To assess the prospects for replication, data collection should focus on testing whether the service is reaching its intended beneficiaries and the fidelity of the service's delivery (the extent to which delivery is faithful to the original model).[25]

In reality, it is likely that some parts of your theory of change will be better evidenced than others. So the evidence gathering process may be more gradual and cyclical than sequential. Knowing which areas of your theory of change are well or poorly evidenced will help you prioritise your research. You can test the less well-evidenced assumptions, while taking a lighter-touch approach for areas with a stronger evidence base.

[‡] Document types which are not controlled by commercial publishers eg, reports, conference proceedings, working papers etc.

[§] A systematic literature review aims to find as much research which is relevant to a particular topic as possible, applying specific methods to determine what should be included and to review the credibility of the evidence provided.

Building an evidence base for poorly-evidenced models

- **Genuinely new models or a poorly-evidenced programmes will require you to start from first principles and gradually build the evidence base.**

This requires more significant investment in research and evaluation. To be able to fully understand how and in what context your programme works (and adapt services on this basis) and to gain credibility in the eyes of external stakeholders you will need to demonstrate effectiveness of your model to a relatively high standard of evidence.

- **Don't just jump to an RCT.**

Evaluating innovative or poorly-evidenced models requires a staged process in which key questions are addressed in turn. The first step may not even focus on evaluating services. Rather, research could assess levels of need in the areas that you are working, and develop insights about how the model could contribute to change. This is crucial for understanding how your programme can best support its intended beneficiaries, and answering the fundamental question of whether it is needed at all.

- **New programmes should start with developmental and formative approaches.**

These evaluation approaches involve iterative testing and quick feedback loops which can support the adaptation of services as information becomes available. This can be based on relatively simple methods of data collection such as user satisfaction, although as the testing continues some evidence on early outcomes should also influence service refinement.

- **Established, but poorly-evidenced, programmes should consider more comprehensive evaluative enquiry.**

This research should focus on explanatory analysis, interrogating the mechanisms of change that support outcomes,[27] as well as demonstrating the outcomes that have been achieved.

Building an evidence base: TheHorseCourse

TheHorseCourse (THC) is a charity that uses an equine-assisted intervention to promote desistance among high-risk, disengaged and disruptive offenders. THC uses a unique approach in the UK criminal justice sector, and the wider evidence that exists on equine-assisted practice consists predominantly of anecdotal reports and case studies. As a result THC has had to work hard at developing the evidence base for its intervention and gaining credibility among key stakeholders in the sector.[26] This has involved receiving support from external evaluators to gradually build the case using a range of evaluation approaches (over a number of studies), including survey data, exit interviews, observation, contribution analysis and a quasi-experimental study.

Framing your research questions

On the guidance table: column 2 shows typical research questions associated with evidence needs.

The considerations above should shape the questions that your evaluation activity seeks to address, giving you a clear sense of what you need to know and what questions can reasonably be answered at that time. These can cover descriptive (what is happening); explanatory (how does it work), or causal questions (what change is due to the programme). As well as research questions relating to the programme or approach, it is also worth considering learning questions more broadly, so your evaluation answers the direct question: *what have you learned*?[28] This can be a useful counterpart to evaluations focused on demonstrating results, helping to ensure that the research design and conduct ultimately helps the organisation to be good at what they do, not just look good, and that this can be shared with others.

Balance evidence needs with available resources

On the guidance table: the recommended methods include options depending on level of resource, to address evidence need.

Evaluation guidance setting out the ideal design and methods to answer particular research questions is difficult to apply given the realities facing the vast majority of social sector organisations. Budgets are tight and organisations are keen to maximise the amount of resource going to directly supporting beneficiaries. It is important therefore to balance the need to produce sufficient, credible evidence—that can meaningfully inform programme delivery and engage stakeholders—with what is realistic given your resources.

Your evidence needs: what kinds of evidence are important?

On the guidance table: column 1 covers evidence need alongside existing evidence available. These are separate but related considerations.

- **What do you need to know and who needs to know it?**

A good place to start is to think about your internal evidence needs (around learning and improving), and the external evidence needs relating to your stakeholders and the sector in which you work (about accountability and demonstrating effectiveness). Evaluation should as far as possible serve both internal and external needs. There can be trade-offs between the resources required to carry out evaluation work and the value this brings to the organisation, or between meeting external and internal needs. INTRAC's recommendation to INGOs is that these are considered on a case by case basis.[29]

- **Think about the kind of information that you need to help improve your services.**

This means thinking about the unanswered how and why questions about your programme or service. It is also important to consider how you would actually use the data you collect, from informing small changes in delivery to much wider strategy decisions. This may not require a particularly complex research design or very high standard outcomes data if, for instance, output data can influence what you do. For example, if you're running a series of events or training sessions, just looking at factors such as attendance and engagement (particularly over time) can help you think about what makes one session work better than another.

- **Identify your key stakeholders and understand their evidence needs.**

Particular stakeholders will have different priorities for the kinds of questions they want answers to and different views about what counts as good evidence. Discussing evidence needs with your stakeholders can increase the efficiency of your evaluative activity and prevent you from collecting data that isn't needed or useful. As part of this, funders and service deliverers may need to communicate and negotiate on evaluation plans, based on existing evidence and mutually agreed priority outcomes. While no stakeholder group will be totally predictable, typical groups and what they often value are suggested below:

Employees, particularly frontline staff, may see the benefits of a service or programme in their day-to-day work. They may need little convincing that their work makes a difference, while data collection may be seen as a burden. Evaluative research should aim to provide useful learning for your staff, though this may vary depending on how well you understand your programmes already (as discussed above).

Beneficiaries' perspectives on services are crucial to evaluating them, therefore the role of evidence to ensure accountability to beneficiaries should be a core aim of any evaluative work. While the type of evidence may be of less significance to beneficiaries, research findings are important to enable them to hold organisations to account, as well as providing guidance on how others have achieved their goals. Beneficiaries can be overlooked when thinking about stakeholders' evidence requirements, but this is a critical stakeholder group.

Commissioners are facing a host of challenges, particularly in balancing shrinking budgets and rising need. Demonstrating that your work is supporting outcomes that align with their strategic priorities is key, but the required standard of evidence around programme efficacy may differ from that of policymakers. Commissioners

are likely to be particularly interested in how these outcomes can be supported within the cost envelope of specific contracts. Hence, as NPC's report into commissioning in arts and cultural sector points out: *'Measuring the social value that activities deliver is crucial as commissioners have to justify their investments and demonstrate that the money spent gets results'.*[30]

Grant-makers and donors' evidence requirements will depend on the kinds of organisation they're looking to fund and the outcomes that they're hoping to support. Sector expectations are important here as certain approaches and metrics may be commonly recognised and valued which align with funder interests. While strong experimentally derived data might attract some funders, many will be just as interested in understanding programme theory and evidence of how change mechanisms might work (which could include qualitative data on service user outcomes).

Many funders will look to work with grantees to develop a mutually beneficial approach to measurement and reporting. If grantees have a good understanding of their evidence needs and a plan for meeting them, then there can be significant scope for influencing reporting criteria. It is important to have realistic plans for outcomes measurement from the outset. NPC's research into funder impact practice found that evidence of impact played a bigger role in decisions to renew funding than awarding initial grants. [31]

Policymakers will often be interested in evaluations that help answer questions related to programme efficacy ('what works'), and will be influenced by government guidance prioritising quantitative data derived from a high standard (counterfactual) research design[33]. Depending on the stage of policy development, however, policymakers will be interested in promising findings and also in answering broader questions, such as *'understanding behaviour change, value for money, unintended outcomes and longer term impacts*'.[34] If you're looking to influence policymakers, how you use your evidence may be just as important as the kind of evidence you produce.[35]

Aligning evidence with sector standards: KL Felicitas Foundation

The KL Felicitas Foundation (KLF) is a pioneer of 100% impact investing, committing 100% of its assets to positive social and environmental impact. In evaluating and measuring its impact KLF has sought to align to (and in the process promote) an emerging standard for measurement within impact investing—IRIS metrics (Impact Reporting & Investment Standards). Applying IRIS metrics to its portfolio enables investees and investors to begin to develop a common understanding about key outputs and outcomes, and to share ideas on investment practice.[32]

Your resources: what's feasible?

Even if you've thought carefully about the specific questions needed to meet your evidence and stakeholder requirements, this will still leave a variety of evaluative approaches which will be more or less resource intensive. You therefore need to consider which approaches best align to your available resources, including your organisation's capacity and capabilities.

- **Assess your capacity.**

This relates to your financial resources and the amount of time you are able to devote to evaluative activity. This should be influenced by a couple of considerations. First, what constitutes 'proportionate' will depend on the kind of evaluative question you're trying to answer. For instance, when piloting and testing a new model you may need to spend a far larger proportion of the overall budget on research to guide that development. Conversely, if you're delivering a model backed by a strong body of evidence, a well-integrated performance monitoring system may mean you can spend relatively little on evaluation.

Second and more broadly, the evaluation budget should be proportionate to the total resources allocated to a programme or service. A popular ballpark recommendation is for 10% of a programme's budget to go towards its evaluation, which is much more than most charities spend.[36] However, budgets should vary widely according to the research need[37]. Testing a prototype intervention could require 20% of an organisation's budget, though monitoring a well-established programme may be closer to 5%.

- **Acknowledge your capabilities.**

This relates to your internal evaluation/research skills, or access to external support. Your chosen approach must match staff's ability to conduct research and make sense of the data or be helped to do so. Introducing methods which don't meet employees' skills risks both undermining the credibility of the results, and weakening staff engagement in evaluation. It may be that you need to seek external advice or commission externally, thus incurring financial costs which relate back to your capacity.

It's important to note that a lack of expert specialism and limited financial resources do not necessarily mean good quality evidence is out of reach. Organisations such as NPC and Money Advice Service, who are interested in supporting more evidence-based practice, have developed tools to help overcome these barriers and support organisations to do quality evaluation.

- **Sharing approaches to measurement can pool skills and resources across networks.**

This can be a more attractive and realistic option than pursuing measurement independently. It may be particularly relevant to smaller organisations that lack the resources to develop their own outcomes measurement framework. For example, Insights is a national outcomes measurement system for domestic abuse organisations in the UK, run by SafeLives. The system measures outcomes indicating victims' safety before and after an intervention. Currently 42 organisations are using Insights, ranging from small, local voluntary organisations to large national charities.[38]

Examples of free tools to support evaluation

- Inspiring Impact: A programme to support improved impact practice within the UK charity sector, run by a consortium of sector bodies and evaluation specialist including NPC. It provides a searchable database of measurement tools and an online step by step self-assessment to help you understand and improve your organisation's impact practice. http://inspiringimpact.org/
- Centre for Youth Impact—Resource Hub: An online information hub to help youth charities navigate the wide range of tools, resources, reports and guidance available on impact and evaluation. www.youthimpact.uk
- Justice Data Lab: An initiative set up by NPC and the Ministry of Justice to enable charities working with offenders to access re-offending data and conduct quasi-experimental analysis comparing their cohort to a matched group of offenders. NPC is looking to set up similar initiatives in education, employment and health. www.NPCdatalabs.org
- Money Advice Service Financial Capability Evaluation Toolkit: An online set of evaluation resources to help organisations delivering programmes that seek to improve people's financial capability. It includes four outcomes frameworks, with indicators and beneficiary survey questions for interventions targeting adults, children and young people, teachers, and parents. www.fincap.org.uk/evaluation-toolkit-homepage

Tailor evaluation to the nature of your intervention

In the guidance table: column 3 shows methodologies for simple interventions; column 4, for more complex.

Acknowledging the complexity of your intervention

- **Understand the 'measurability' of your service or programme.**

For instance, you might want to ask questions about programme efficacy but are working in a context that makes setting up comparison groups impossible. This doesn't mean you should abandon trying to find answers to your research question, but there will be other more appropriate methods for doing so. Placing interventions on a spectrum from simple to complex can help in thinking through appropriate evaluation methods.[39]

- **'Simple' interventions lend themselves well to experimental evaluation design.**

These are interventions where there is a clear relationship between delivered services and beneficiary outcomes, and few intervening factors to consider. 'Simple' interventions are those with a tightly defined service or product being delivered by a single organisation. They target few outcomes, with hard indicators, for example qualifications or re-offending rates, which lend themselves to measurement and monitoring. Experimental study designs can work well with these types of interventions as it should be reasonably straightforward to create a comparison group, if changes occur within relatively short timeframes.

- **'Complex' interventions require a greater triangulation of methods.**

These are interventions where there may be no primary cause and effect relationship, the overall programme could be made up of multiple and diverse services or products, and multiple stakeholders change over time. The outcomes targeted by the programme may also be more challenging to measure, and focus on softer outcomes such as attitudinal or behaviour change.

In these contexts it is rarely possible to create a comparison group, and the evaluation design may have to flex with changing circumstances. Experimental or solely outcomes approaches will not deliver meaningful results—without a tightly defined programme, it is not possible to attribute changes to one set of activities with a high degree of certainty. The focus should therefore be on triangulation, using a range of different and appropriate methodologies to gauge effects. Knowledge can be cumulatively built about the extent to which change is happening and to identify the mechanisms bringing this about.

- **Distance from the beneficiary is another key variable to consider**.

The challenges of complex frontline delivery settings are amplified if organisations are working through others and are one or more steps removed from the ultimate beneficiary. For funders, social investors and other intermediary organisations isolating their impact can be a substantial challenge, particularly if they only contributed a small proportion of the funding or work. As well as influencing the appropriate evaluative method, it is important to be realistic about the kinds of question you can meaningfully and credibly answer in cases like these.

- **Assessing contribution may be more meaningful that attribution in complex settings.**

In international development it has, for instance, become increasingly accepted that trying to attribute outcomes to a particular funder or NGO working with multiple 'on-the-ground' partners may not be possible or worthwhile. As Bond's guide to impact evaluation notes: '*identifying your contribution and recognising the contribution of others is more realistic than searching for evidence of sole attribution*'. The next page provides examples of how organisations can account for the simplicity/complexity of an intervention when evaluating their work.

Taking a tailored approach for different types of intervention/contexts

Talent Ed: evaluating learning support and mentoring for talented students from disadvantaged backgrounds.

TalentEd was set up in 2012 and is run by a small group of staff. Despite limited resources it has been able to develop a rigorous approach to monitoring and evaluation, through focusing on understanding its impact on hard outcomes related to educational attainment. To do this in a cost effective way it has drawn on the existing data within the school/education system to monitor outcomes of its service users and compare to national/regional averages.

Imara: evaluating early intervention with victims of child sexual abuse.[40]

Imara has encountered a number of challenges when measuring the impact of interventions supporting victims of child sexual abuse. This has included understanding overall progress despite major fluctuations in emotions, controlling for intervening variables (eg, the impact of the criminal justice process), and understanding the long-term effects of the intervention. These challenges mean that quantitative measures have been less useful to develop an accurate understanding of impact. As a result so far Imara has focused on using qualitative approaches when evaluating its work, to gain a deep understanding of the mechanisms of change and the contexts in which its activities are most effective. It has also used quantitative analysis of trial outcomes.

Oxfam: evaluating advocacy campaigns.[41]

Oxfam's approach to evaluating its advocacy campaigns has developed out of a need to understand its impact in a context in which it is impossible to implement experimental and statistical approaches. As a result Oxfam has developed an evaluative approach based on a methodology known as process-tracing. As part of this an external evaluator randomly selects around 8 campaigns annually and looks for evidence that can link the project to the intend outcomes. Alongside this, evidence is also sought for alternative 'stories' of how these changes came about, which is compared to those related to Oxfam's work. This comparison enables Oxfam to develop an understanding of the level of its contribution to the observed outcomes.

CHOOSING AN EVALUATION DESIGN

Guidance on methodological considerations

Thinking through the considerations outlined helps determine what type of evaluative research is appropriate to you. While it is impossible to definitively suggest any evaluation approach without an understanding of an organisation's individual situation, it is possible to lay out the types of methodologies that are generally well suited to different situations. To this end, the guidance table in this section shows how these considerations can play out. It covers the evidence you may need to design, develop and assess programmes, which goes beyond formal evaluative activity. This includes evidence necessary to inform and design a programme, as well as evaluate it.

Using the guidance table

The table outlines how the evidence available about a programme/approach, combined with what you and your stakeholders want to use the evidence for, will determine what your research questions are. Research questions direct the approach to evaluation, with a choice of methodological options depending on the nature and complexity of the intervention and the resources available to you. This is not an exhaustive list, but offers a sense of the types of evidence you might consider to address your questions and accommodate your constraints.

Use columns 1 and 2 to identify the evidence needs that match your situation. Ask:

- What is the strength of the evidence available for my programme/approach?
- What do we and our stakeholders want to use the evidence for?

This gives you your evidence needs, and thus the **research questions** you will have.
Then use columns 3 and 4 to find potential methods that match your intervention and resources. Ask:

- How complex is the programme/approach?
- What resources can be made available?

This suggests which **methodologies** may provide proportionate evaluation approaches.

Organisations will not necessarily move through the different types of evaluation activity described—there is no linear path in terms of the evidence bases that organisations must travel through. The table guides the user with two considerations: the level of evidence you have available, and the evidence requirements of your situation. This also aligns with the stage of development of an initiative (from innovation through to replication). However, your use of methods across the table is not necessarily a linear journey. You may have been implementing a programme for some time, but without any evidence for it. Using mixed methods evaluation could show the changes and improvements that can be made, while providing useful evidence to others on what the programme mechanisms are. This particular method may prove useful for a long time, to repeatedly inform improvements.

The approaches provided include designs across a spectrum of complexity. Some of these certainly require evaluation expertise and are likely to involve external commissioning, though many could be conducted in-house. The decision about who should conduct the work will be based on an assessment of internal capacity and capability and the relative value of an external, independent voice. If conducting the evaluation internally, conflict of interest issues will always need to be considered.

Guidance table

<table>
<tr><th>1
Evidence base and evidence need</th><th>2
Research questions</th><th colspan="2">3
Simple programme</th><th>4
Complex set of activities or interdependencies</th></tr>
<tr><td>1. Unknown</td><td>• What's the evidence?</td><td colspan="3">Undertake a scoping review to map out any existing evidence on the topic.</td></tr>
<tr><td rowspan="2">2. Low or limited evidence. There are no known effective interventions in this area, or they are not found in comparable contexts. OR the intervention is in use, but not evidenced.</td><td rowspan="2">• What changes are sought?
• What are the drivers and barriers to change?
• What sort of intervention is needed?</td><td colspan="3">Needs analysis (descriptive questions)</td></tr>
<tr><td colspan="3">Review current practice in other settings.
Use desk research to build a picture of the issue.
Consider qualitative research with potential service users. Consider stakeholder research to understand how the issue interacts with other areas or dimensions of the problem.</td></tr>
<tr><td rowspan="3">3. An approach to delivering an intervention is being planned, based on needs analysis. Some links to evidence support the approach , but do not demonstrate its efficacy. OR the intervention is in use, but not well evidenced.</td><td rowspan="3">• How should/does the intervention work?
• What should make it work?
• How will we know if it's working?</td><td colspan="3">Identifying key mechanisms for change (theorised causal and explanatory questions)</td></tr>
<tr><td colspan="2">Develop a theory of change, using existing evidence where possible. Specify outcome measures and priorities for measurement.</td><td>See column 3. In addition, map roles of other stakeholders and factors influencing the programme. Consider if their data could be used.</td></tr>
<tr><td colspan="3">Test the assumptions in the theory of change. Revise and refine it as much/often as needed.</td></tr>
<tr><td rowspan="2">4. A new approach is being tested, with little evidence or analysis to support it. There is a need for feedback to support decisions and improve the approach.</td><td rowspan="2">• What are the characteristics of the emerging service?
• Are staff/users engaging?
• Are there any teething problems with delivery?
• What do early indicators suggest about progress?</td><td colspan="3">Early dev & testing (descriptive and evaluative questions): developmental evaluation</td></tr>
<tr><td colspan="2">Select simple indicators for monitoring and rapid review. Use mixed methods data where possible, including qualitative interviews with staff and users on unintended results.
With more resource: Create more precise, short-term questions to trial different programme components. Collect baseline data for later pre and post comparison.</td><td>See column 3. In addition, seek feedback from stakeholders who influence the programme.
With more resource: use a greater variety of methods eg, observation, groups, and monitoring information.</td></tr>
</table>

<table>
<tr><th>1
Evidence base and evidence need</th><th>2
Research questions</th><th>3
Simple programme</th><th>4
Complex set of activities or interdependencies</th></tr>
<tr><td rowspan="2">5. A programme/approach uses existing evidence to support assumptions with evidence of promise, but a more rigorous examination is needed to inform decisions and improvements.</td><td rowspan="2">How does the programme work/what are the mechanisms for change?
Are indicators positive?
What's working well/less well?
In which circumstances does this work?
For whom does it work?
What changes could make the programme more effective?
Could it be more targeted?</td><td colspan="2">Understanding, refining and improving an approach (descriptive and explanatory and some causal questions): process evaluation</td></tr>
<tr><td>Develop and revise a theory of change to tightly reflect programme mechanisms. Use qualitative staff interviews to gauge the link between programme and outcomes.

With more resource: Process evaluation may be appropriate. Use mixed methods such as interviews, observations and focus groups, quantitative monitoring information and pre/post early outcomes data.</td><td>See column 3.

Limitations: Interactions with context may be complex; causal pathways are hard to establish.

With more resource: Explore more complex mechanisms, by comparing beneficiary groups and different settings. Focus on seeking patterns in outcomes and context to identify conditions the intervention will or won't work in, and for whom.</td></tr>
<tr><td rowspan="2">6. There is a need to rigorously demonstrate a causal link between the programme/approach and the outcomes and measure the extent of change, with little existing evidence.</td><td rowspan="2">Does the intervention work?
Did the activities lead to the changes seen; are changes due to the programme?
What makes the approach work?</td><td colspan="2">Demonstrating changes are due to the programme: impact evaluation</td></tr>
<tr><td>Consider contribution analysis with a comparison group if feasible. Collect pre and post data. Explore alternative explanations for changes with MI and qualitative research.

With more resource: Experimental or quasi-experimental approaches may be appropriate to test a theoretical model. Consider a Randomised Control Trial. If not feasible, consider statistical techniques to construct a comparison group.
Limitations of experimental approach: Unintended impacts are not explored. For a very small scale intervention, may not be feasible.</td><td>Use multiple data sources to strengthen contribution analysis. Consider user surveys, case studies and case-based comparisons between settings. Secondary data analysis on correlated changes may be useful. Explore other factors that could have contributed, eg, using stakeholder interviews or monitoring information.

With more resource: Detailed theory-driven evaluation may reveal multiple or complex pathways. Compare cases within and across settings to identify different pathways.

Consider Qualitative Comparative Analysis to analyse causal contributions, with external support

Limitations: Not viable for very complex areas.</td></tr>
</table>

<table>
<tr><th>1
Evidence base and evidence need</th><th>2
Research questions</th><th>3
Simple programme</th><th>4
Complex set of activities or interdependencies</th></tr>
<tr><td rowspan="2">7. There is a need to decide whether the programme/approach should be supported, based on a variety of existing evidence sources.</td><td rowspan="2">On balance, is this approach achieving goals? Is it worth continuing?
Is it good value compared to other approaches?
Are the benefits worth the costs?</td><td colspan="2">Consider Lessons Learnt analysis: eg, synthesise outcomes data, user and stakeholder feedback and evidence on alternative approaches. Frame by the questions to be answered.</td></tr>
<tr><td>With more resource: Consider economic evaluation: cost-benefit analysis, cost-effectiveness analysis or Social Return On Investment analysis. Explore unintended impacts using qualitative and monitoring information.</td><td>With more resource: It is challenging to provide a financial value for complex programmes. Consider stakeholder and user consultation and unintended impacts analysis (see column 3).</td></tr>
<tr><td>8. The programme is grounded in good evidence and fits the setting. Delivery and accountability are priorities.</td><td>Is the programme being implemented properly?
Are users engaged?
Are the expected outcomes being achieved?</td><td colspan="2">Performance monitoring
Collect light-touch survey or qualitative data from a sample of service users.
Performance manage staff to ensure fidelity to the programme approach.
Limitation: this will not pick up emergent unintended consequences if the setting is changing.
With more resource: Multi-stakeholder research and/or monitoring information could also be used to monitor the operating context.</td></tr>
<tr><td rowspan="2">9. The programme/approach is demonstrated to be effective but evaluation is limited to one setting, and it is unknown whether and how it could be replicated or scaled.</td><td rowspan="2">Can the approach be used in the new setting?
What needs to be tailored, and how?
Are results consistent?</td><td colspan="2">Research to support scaling and adaptation
Consider a rapid review of evidence and an analysis of local need to identify likely differences in setting. Pilot the approach with regular reviews, using qualitative data and building up to short-term quantitative intermediate outcomes.</td></tr>
<tr><td>With more resource: Consider more systematic review to explore whether these show similar effects.</td><td>With more resource: Use formative testing of adaptations, to explore context and mechanisms for different programme components.</td></tr>
</table>

FINAL CONSIDERATIONS

Evaluations serve a range of purposes and generate different types of evidence to do so. Full-scale impact evaluations have been shown to be inappropriate or disproportionate in many situations, but broader evaluative activities suited to the organisation's needs should help both large and small decisions, while also cultivating an impact mind-set.

An impact mind-set

As an organisation, we suggest you seek to articulate the impact you want to achieve, for whom, and develop a robust model that shows how you will deliver this. This approach should be understood by your staff and volunteers and should be supported by academic evidence and good practice in the field in which you work. Focusing on your intended outcomes, and using research practices to support learning and improvement, should be the foundations for effective services for beneficiaries, and support the development of the wider sector.

Further resources

NPC's *four pillar approach* guides you through the key steps in developing an impact measurement framework to help you understand the difference you make. www.thinknpc.org/publications/npcs-four-pillar-approach/

NPC's practical guide to creating your theory of change helps you think through the process from beginning to end, including how to represent and use it. www.thinknpc.org/publications/creating-your-theory-of-change/

NPC's *Listen and learn* explains how charities can use qualitative research to inform their decisions and help identify what works. www.thinknpc.org/publications/listen-and-learn-qualitative-research/

This paper has been written as part of the Money Advice Service Financial Capability Evaluation Toolkit, an online set of evaluation resources to help organisations delivering programmes that seek to improve people's financial capability. www.fincap.org.uk/evaluation-toolkit-homepage

Read more about NPC's impact measurement services here and look out for our upcoming seminars and training events.

If you'd like to discuss your measurement approach, or require further information about our services, please get in touch via info@thinknpc.org.

REFERENCES

[1] Richard Clegg quoted in Corfe, E. (2016) 'Impact measurement risks becoming a 'cottage industry', says Lloyd's Register Foundation director', in *Civil Society Magazine*, 18 January 2016.

[2] Nutley, S., Powell, A. and Davies, H. (2013) *What Counts as Good Evidence?* Alliance for Useful Evidence.

[3] See Caroline Fiennes and Ken Berger (2016) 'Oops: we made the non-profit impact revolution go wrong', in *Alliance Magazine*, 15 March 2016.

[4] Kazimirski, A. and Pritchard, D. (2014) *Building your Measurement Framework: NPC's four pillar approach*. New Philanthropy Capital.

[5] Hestbaek, C. (2014) *Closing in on change: Measuring the effectiveness of your campaign*. New Philanthropy Capital.

[6] https://www.gov.uk/guidance/what-works-network

[7] Pawson, R. and Tilley, N. (1997) *Realistic evaluation*. Sage Publications.

[8] Pawson, R. (2006) *Evidence-based policy: a realist perspective*. Sage Publications.

[9] Realising Ambition (2016) *Evidence is Confidence: How to create a richer evidential tapestry*. Realising Ambition Programme Insights: Issue 2.

[10] Axford, N. (2016) *Is 'What works?' a question we should still be seeking to answer?*, Centre for Youth Impact.

[11] Stern, E (2015) *Impact evaluation: A guide for commissioners and managers*. Bond.

[12] Puttick, R. and Ludlow, J. (2013) *Standards of evidence: an approach that balances the need for evidence with innovation*. Nesta.

[13] Education Endowment Foundation (2014) *Classification of the security of findings from EEF evaluations*.

[14] Bull, D., Bagwell, S., Joy, I. and Weston, A. (2016) *Untapped Potential: Bringing the voluntary sector's strengths to health and care transformation*. New Philanthropy Capital.

[15] Boaz, A. and Ashby, D. (2003) *Fit for purpose? Assessing research quality for evidence based policy and practice*, ESRC UK Centre for Evidence Based Policy and Practice: Working Paper 11.

[16] Ginsburg, A. and Smith, M.S. (2016) *Do randomized control trials meet the 'gold standard'?* American Enterprise Institute.

[17] Bond's evidence principles: https://www.bond.org.uk/data/files/Effectiveness_Programme/120828Full_Bond_checklist_and_guide.pdf

[18] Breckon, J. (2016) *Using research evidence: a practice guide*. Nesta

[19] Handley, S., Sabri, F. and Kazimirski, A (2016) *Shared measurement: Greater than the sum of its parts*. Inspiring Impact

[20] Bagwell, S., Hestbaek, C., Harries, E. and Kail, A. (2014) *Financial capability outcome frameworks*. New Philanthropy Capital and the Money Advice Service.

[21] Pawson, R. *op cit*

[22] Kail, A, and Lumley, T. (2012) *Theory of change: The beginning of making a difference*. New Philanthropy Capital.

[23] Kazimirski, A., Bagwell, S. and Gripper, R. (2016) *NHS Charities: Evaluation guidance*. New Philanthropy Capital.

[24] Kazimirski, A. and Pritchard, D. (2014) *Building your Measurement Framework: NPC's four pillar approach*. New Philanthropy Capital.

[25] Realising Ambition (2016) *Evidence is Confidence: How to create a richer evidential tapestry.* Realising Ambition Programme Insights: Issue 2.

[26] TheHorseCourse (2016) *Summary of evidence to date.*

[27] Pawson, R. *op cit.*

[28] Simister, N. (2016) *Summarising portfolio change: results framework at organisational level.* INTRAC M&E Paper 10.

[29] Simister, N. *op cit.*

[30] Bagwell, D., Bull, D., Joy, I. and Svistak, M. (2014) *Opportunities for alignment: arts and cultural organisations and public sector commissioning*. New Philanthropy Capital.

[31] Kail, A., Van Vliet, A. and Baumgartner, L. (2012) *Funding Impact: impact measurement practices among funders in the UK*. New Philanthropy Capital.

[32] Harrison-Evans, P., Lomax, P. and Rotheroe, A. (2015) *Investing for impact: Practical tools, lessons, and results*. New Philanthropy Capital.

[33] HM Treasury (2011) *The Magenta Book: Guidance for evaluation*.

[34] Hills, D and Junge, K. (2010) *Guidance for transport impact evaluations: Choosing an evaluation approach to achieve better attribution*. Tavistock Institute.

[35] Breckon, J. and Dodson, J. (2016) *Using evidence: what works?* Alliance for Useful Evidence.

[36] Ni Ogain, E., Lumley T. and Pritchard, D. (2012) *Making an Impact: Impact measurement among charities and social enterprises in the UK.* New Philanthropy Capital.

[37] Pritchard, D. (2012) 'How much evidence is enough?', *NPC blog*, 30 October 2012.

[38] http://www.caada.org.uk/practice-support/resources-domestic-abuse-and-idva-service-managers/insights

[39] Rogers, P., Hawkins, A., McDonald, B., Macfarlen, A. and Milne, C. (2015) *Choosing appropriate designs and methods for impact evaluation*. Australian Government: Department of Industry, Innovation and Science.

[40] Svistak, M. and Kazimirski, A. (2015) *Support after trauma: Assessing the impact of Imara, a child sexual abuse (CSA) support programme in Nottingham*. New Philanthropy Capital.

[41] Hutchings, C. and Bowman, K. (2013) 'Lessons learned from a year of advocacy impact evaluation', *Oxfam Policy and Practice Blog*, 28 February 2013

TRANSFORMING THE CHARITY SECTOR

NPC is a charity think tank and consultancy which occupies a unique position at the nexus between charities and funders, helping them achieve the greatest impact. We are driven by the values and mission of the charity sector, to which we bring the rigour, clarity and analysis needed to better achieve the outcomes we all seek. We also share the motivations and passion of funders, to which we bring our expertise, experience and track record of success.

Increasing the impact of charities: NPC exists to make charities and social enterprises more successful in achieving their missions. Through rigorous analysis, practical advice and innovative thinking, we make charities' money and energy go further, and help them to achieve the greatest impact.

Increasing the impact of funders: NPC's role is to make funders more successful too. We share the passion funders have for helping charities and changing people's lives. We understand their motivations and their objectives, and we know that giving is more rewarding if it achieves the greatest impact it can.

Strengthening the partnership between charities and funders: NPC's mission is also to bring the two sides of the funding equation together, improving understanding and enhancing their combined impact. We can help funders and those they fund to connect and transform the way they work together to achieve their vision.

New Philanthropy Capital
185 Park Street, London SE1 9BL
020 7620 4850
info@thinkNPC.org

Registered charity No 1091450
A company limited by guarantee
Registered in England and Wales No 4244715

www.thinkNPC.org