BUILDING YOUR MEASUREMENT FRAMEWORK: NPC'S FOUR PILLAR APPROACH

Anne Kazimirski and David Pritchard

June 2014

Acknowledgments

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 developed by others.

Too many individuals and organisations have influenced this guide in one way or another to name them all individually, so we can only offer a blanket acknowledgement of their contribution.

However, it is important to mention the support and guidance we have received from our colleagues at NPC—particularly Lucy de Las Casas, Eibhlín Ní Ógáin, James Noble, Marina Svistak and Alex van Vliet—and from external reviewer David Hunter.

We would also like to thank one group which has been key to the development of our approach: our clients. As of April 2014, we have completed consulting projects for 16 organisations using and testing our approach.

We are particularly grateful to the case studies included in this guidance: Business in the Community, Send a Cow, Shaw Trust, and the Women's Counselling and Therapy Service.

FOREWORD

Why NPC is publishing its approach to measurement

Few join the charity sector for the sake of evaluation, measurement or statistics. But charities must engage with these activities to get a proper feel for how effective they are at delivering for the beneficiaries and causes they care about. With many operating in an increasingly competitive environment and under mounting pressure from funders and commissioners to demonstrate their social impact, it is an agenda that cannot be ignored.

But many wanting to begin on this journey find it difficult to know where to start.

It is hard to navigate the choices that surround what to measure and how often, which methodology and tools to use under what circumstance, and how much time and money to spend on the whole exercise.

At NPC, we guide charities through this maze by helping them make decisions that work best for their organisation and cause. We are a mission-driven organisation ourselves, aiming to transform the charity sector so it continues to do better.

As part of this, we want everyone to have access to a measurement framework—precisely what this guidance is about. Our four pillar approach, presented here, sits at the heart of what we do. It is the same approach we use with clients, and one that has been tried and tested over many years' experience.

Of course, context is important. We cannot address all eventualities here and questions will undoubtedly remain—our experts and consultants stand ready to help.

Dan Corry

Chief Executive, NPC

CONTENTS

INTRODUCTION

At NPC we believe in impact measurement as a way for charities and funders to increase their effectiveness. We promote the benefits of good impact measurement across the charity sector and encourage organisations to overcome the inherent challenges. We do this so that organisations can improve what they do and deliver the best results for their beneficiaries.

We champion impact measurement through our published advice on developing an impact-led approach, our conferences on the topic, and our consulting work and training sessions with a wide range of charities and funders. We have experienced every stage of the impact measurement journey—from encouraging buy-in among senior staff and trustees and identifying what to measure, to developing tools to collect data and thoroughly analysing the results.

In the last ten years we have seen a shift in attitudes to impact measurement and to measurement practice itself. In 2012, our *Making an impact* research found that just over half of charities measure the impact of nearly all their activities.[1] In 2013, *Funding impact* highlighted the increasingly positive attitude of funders to impact measurement: 88% think that it makes charities more effective.[2] Detailed frameworks have sprung up and rigorous measurement tools exist for the vast majority of social outcomes, but the opportunities offered by this kind of analysis have not been grasped by all.

Of course, measuring impact is not easy and our work has revealed a number of barriers to making progress. There are no universal measures of a charity's impact that mirror share price, profit margins or price earnings ratios in the private sector. Despite many well-intentioned efforts, no-one has developed measures of impact that suit all charities—even the term 'impact' is subject to different interpretations.[3] Add to this the challenge of setting aside precious resource, and other contextual and cultural factors—what is already known about what works and what does not, internal skills and experience, the fickle nature of funder requirements and the size of the window to collect data from beneficiaries, and so on—and it is little wonder that some charities struggle to put impact measurement into practice.

This guidance hopes to further our mission to transform the charity sector by helping your organisation to maximise its impact. We explain the key steps involved in developing or improving a measurement framework, and provide practical support along the way.

About this report

We receive a large number of enquiries about how to approach impact measurement. In response to this demand for clear and practical guidance on the topic, we have decided to publish NPC's approach. We hope it enables you to develop a measurement framework that works for your organisation's individual situation in a way that is sufficiently robust for its needs. Ours is not the only approach, but we know through experience that it works for a variety of sizes and types of charities and funders, and results in an efficient, understandable and useful end product.[*]

This guidance can be applied across an organisation or to individual services or programmes. At each stage in the process we refer to the lessons learned by organisations we have worked with, so you can benefit from their experience. We are open and honest about what has been critical to success, and what has been less useful—

[*] For more information see Appendix 1.

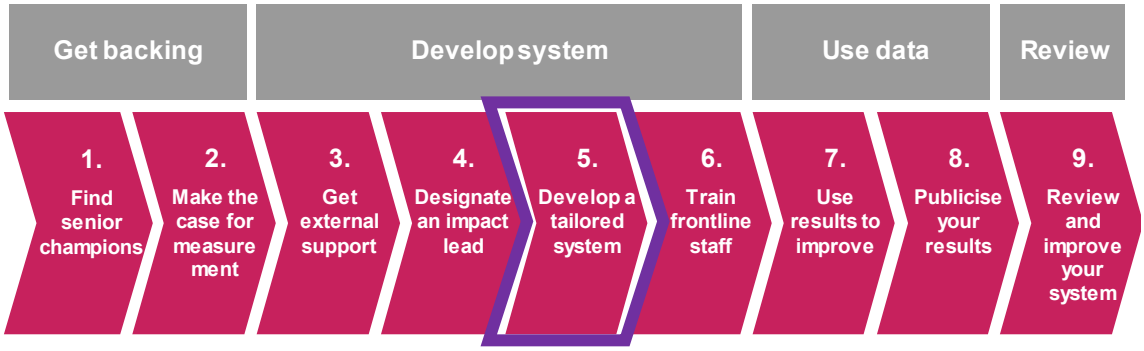
and we are keen to hear from other organisations that have adopted or tried to adopt NPC's approach so we can continue to improve it.

Unfortunately, we cannot answer everything here: organisations have different goals and operate in different environments, and there are too many considerations, options and choices to be able to cover all scenarios. We raise the key questions and guide you to find your own answers. This often leads to the infuriating *'it depends on ',* and you may find you need external support—from NPC or others—to think it through further.

A journey to greater impact

This guidance focuses on developing and implementing an impact measurement framework—an important step in the journey to becoming an impact-driven organisation. It relates to step five of nine—*'Develop a tailored system'*—described in our report *A journey to greater impact.*[4]

Figure 1: A journey to greater impact



Building a tailored system—your own impact measurement framework—cannot be done in isolation. Its success depends on a strong foundation, working through steps one to four, including support for measurement from the top of your organisation, and having the right internal capabilities in place.

Throughout this guidance we talk about the importance of culture and mind-set in developing and implementing a good impact measurement framework. Equally crucial is how you use data to improve your services, and how you report and publicise your results. The development and implementation of a measurement system is not an end in itself: the benefits you see and the value you gain from measuring your impact depend on what you do with the findings. Use them to review and improve your work, and the investment in developing your measurement approach will be more than worthwhile.

Key terms

Social impact: The effect of an activity on the social fabric of the community and well-being of the individuals and families (http://www.businessdictionary.com).

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.

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).

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.

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

Measurement tools: A means of collecting data. Common tools are surveys, interview questions, observation, and case records (of charities and public sector agencies).

Measurement framework: A list of the outcomes that a charity seeks to achieve together with information on: indicators of those outcomes; data sources; methods of analysis; and measures of outputs, processes, and quality.

Impact measurement: The set of practices through which an organisation establishes what difference its work makes. We use the term to define a broad set of practices that includes measuring both outputs and outcomes. This does not mean we equate measuring outputs with measuring impact—it is just part of the overall process.

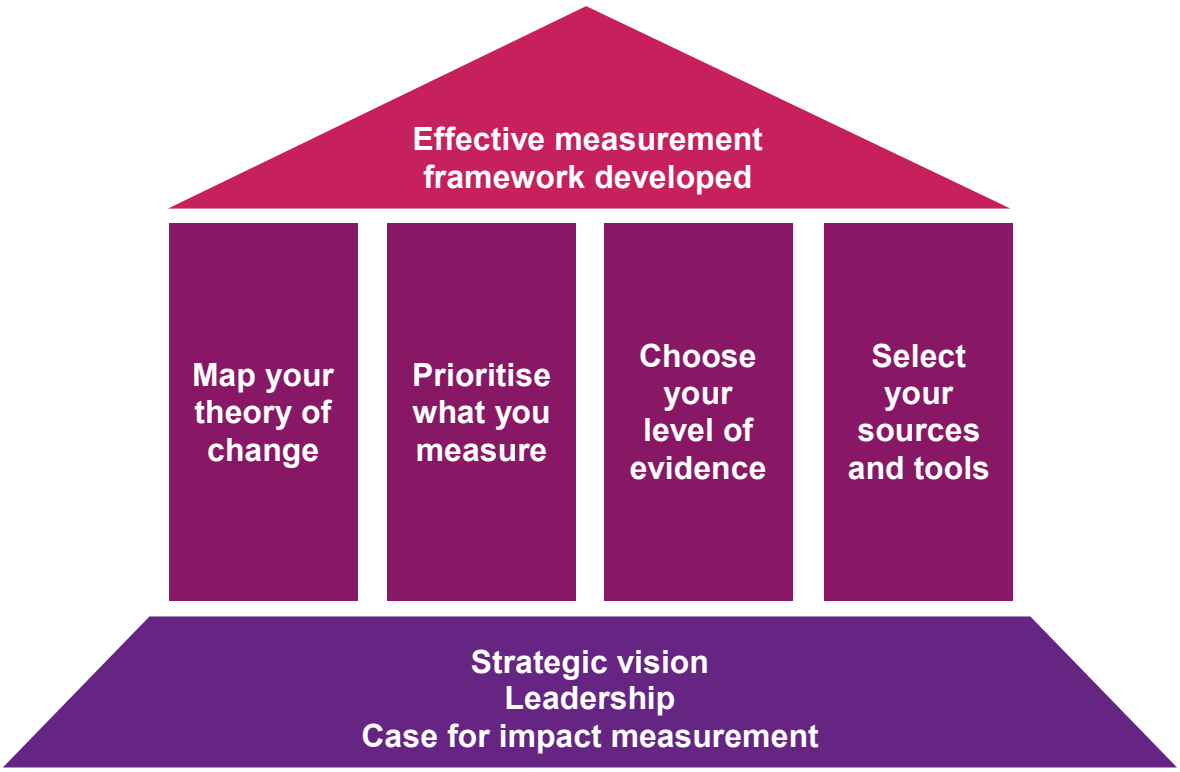
Monitoring: A systematic way to collect and record information to check progress against plans and enable evaluation.

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

A good source of definitions of evaluation terms used by charities can be found at www.jargonbusters.org.uk.

NPC'S FOUR PILLAR APPROACH

Figure 2: NPC's four pillar approach



Building an effective measurement framework

NPC's four pillar approach is a process that your organisation can work through to build an effective measurement framework, ensuring you are measuring the right things in the right way. It aims to provide results that can be used to understand and improve charities' services, as well as report on their progress. The approach is simple in theory but can be tricky to implement well. Mechanically going through the motions is not enough: careful thought and decisions are required to fit each step to your particular organisation.

Building an effective measurement framework relies on several component parts which need to work together—illustrated in Figure 2. A solid foundation is essential, and requires good leadership and buy-in from senior staff and trustees, a strong commitment to the value of impact measurement, and the investment of time, energy and skills. The four pillar approach follows on from here.

Key to developing a successful framework is to distinguish between understanding what you aim to achieve—the change you want to bring about[*]—and how to measure that change. The order is critical. Thinking about how to measure without properly understanding the change you want to bring about will lead to frustration and wasted

[*] The change a charity wants to bring about is the difference it makes compared to the counterfactual. For some charities this might mean trying to prevent things from getting worse rather than making improvements. Other charities may exist to provide relief in times of short-term crisis rather than seeking long-term change.

resources. By being clear at the outset about what you want to measure and why, you are more likely to design, implement and run a measurement system that works for you. The ultimate test of a measurement framework is whether it generates useful data that you can learn from and use as the basis of your reporting. If it does not do this, the framework needs fixing.

Because understanding what you aim to achieve is so important, we begin with a **theory of change**—a map of how an organisation, service or project intends to deliver its desired outcomes. Designing a measurement framework around your theory of change will then ensure you collect information that tells you what difference you are making. This will involve deciding what data to collect, the level of rigour of evidence you need, and how to go about collecting this data.

The four pillars

This guidance will take you through each of the four steps to developing a successful measurement framework, outlined briefly below.

Step one: Map your theory of change

Many charities struggle to know where to begin with impact measurement. Starting with your goals and developing a theory of change provides clarity, revealing the causal links between what you do and what you are trying to achieve. It gives you a coherent framework to underpin your measurement efforts so that you can move away from ad hoc, opportunist data collection.

Step two: Prioritise what you measure

Collecting the right amount of quality data is key, and getting there might require some trial and error. Prioritise the most important outcomes in your theory of change and focus on measuring those. Your impact is likely to be diffuse, affecting different people in different ways over different time frames, but trying to capture all these changes is complicated and may not be the best use of limited resources. Do not be tempted to collect data that is convenient—just because something is easy to measure, does not mean it is important. Remember to consider data on the possible negative unintended consequences of your work too; this will help you to improve your work.

Step three: Choose your level of evidence

Choose an appropriate level of rigour for evidence of your impact that suits the needs of your stakeholders. There is no one size fits all: it will depend on your own needs, resources and capabilities, and those of your audience. Funders may want a certain level of rigour for a project they have funded; academics or government research departments may want to see a counterfactual.

Step four: Select your sources and tools

Decide what data you need and select or develop measurement tools or data sources to capture it. You may find an existing tool or data source, or you may need to develop one. Do not feel the need to reinvent the wheel: consider what tools are already available, and think about existing evidence for the causal links in your theory of change. It is important to use measurement tools that are fit-for-purpose, and that capture the change you want to bring about.

Benefits of our four pillar approach

It is grounded in what you do and want to achieve. Being based on clear goals and a well-developed theory of change, the evidence you collect better reflects the change you want to bring about.

It gets the organisation working together and exposes different perspectives. Developing a theory of change reveals where there are common goals and where there are differences. Ignoring the differences can lead to inefficiency at best, and your charity working against itself at worst.

It integrates the purposes of different services and activities. Embedded in a theory of change, you can measure how different services and activities contribute to your ultimate goal(s).

It will be welcomed by staff. Implementing this approach incorporates the views of staff at all levels into the theory of change and what is measured, so staff will understand what data is being collected and why. This is critical—typically, staff collect the data that goes into the framework and implement any changes driven by the findings. Without their buy-in, a key purpose of your measurement framework—to drive improvement—cannot be realised.

It produces a measurement framework that is understandable. The theory of change provides a narrative about how your charity has an impact, and what you measure can provide quantitative support. This makes it easy for your staff, beneficiaries, funders and other stakeholders to understand what you do and the impact you intend to have.

It produces a measurement framework that is efficient. By focusing on the change you want to bring about and by using existing data and evidence where possible, the approach minimises the time and effort needed to collect useful impact data.

It produces a practical and useful framework. When populated with data, the measurement framework will help managers to identify where improvements can be made. Indeed, if the framework does not, then it is not useful.

STEP ONE: MAP YOUR THEORY OF CHANGE

The benefits of theory of change

We place a theory of change at the heart of our approach to developing a measurement framework.* It shows what you want to achieve and how you plan to achieve it, setting out the causal links between your activities and your end goal. A clear and robust theory of change is a necessary basis for measuring your impact because it provides a theoretical framework to allow you to assess whether what you do is working as planned, and how it can be improved.

Done properly, a theory of change has three major benefits:

- **It helps you understand all the important outcomes of your work,** so you can develop a framework that measures the right things. These might be intermediate outcomes that lead to several others, or outcomes that make your intervention different from the usual practice. If measurement is not based on a theory of change, it risks not measuring the most important things and therefore wasting money. In this way, a theory of change can help you to ensure that what you measure is related to your strategy.
- **It helps you understand how the outcomes you seek to influence are connected.** Charities that base their measurement on a theory of change can understand *how* change is happening as well as *whether* it is happening. Because a theory of change shows what a charity is trying to achieve and how it is planning to get there, charities can work out whether they are on track to achieve intended outcomes.
- **It helps you track the progress you make towards your final goal.** Some charities' final goals cannot be easily measured—they involve change that happens too gradually, or change in the lives of people who are difficult to track. Theories of change show all the intermediate steps that lead to the end goal, and can therefore help charities to assess their progress towards it, and build their case, even if the goal itself cannot be measured.

A theory of change can help your organisation in other ways too. You can use it to develop and refine your strategy, where it helps to keep a clear goal in mind. You can also use it to think about your place in the sector, and identify areas where you are working together with others, and ways you might collaborate to increase your impact. For more information on developing and using a theory of change, we recommend NPC's paper, *Theory of change: The beginning of making a difference*.[5]

Why intermediate outcomes are important

Intermediate outcomes are preconditions you have to meet to reach a final goal. The link between intermediate outcomes and final goals may have already been shown by other research (for example, the link between happiness and success in school) or is plainly logical (attending school and being successful in school). If you can show improvements to intermediate outcomes then it is reasonable to expect the final outcomes to occur. This can reassure stakeholders and funders that the charity is making progress and can help charities to work out what they can attribute to their work. It is particularly useful for charities involved in campaigning or advocacy work.

* Much of this summarises material from NPC's *Introduction to theory of change*, available at: www.thinknpc.org/publications/theory-of-change/

How to develop a theory of change

To get the most out of your measurement framework, it needs to be understood and embraced by everyone involved in your work. Without this buy-in you will find it difficult to get staff to collect high quality data and to use it in a meaningful way, and much of the value of measuring your impact will be lost.

The best way to engage people is to involve them in developing your theory of change. You can do this in a variety of ways: we often use workshops or a series of interviews to help us develop a draft, followed by a workshop to review and revise it. Whichever approach is taken, we recommend the final version is reviewed by a group of representatives from across the organisation.

Reaching some degree of consensus on the content of the theory of change and the measurement framework is critical to success. Trustees, senior management and front-line staff need to be represented and consulted about which outcomes to prioritise, the level of evidence you need, and the best methodology for collecting data. Staff will be able to flag up at once if your approach is realistic and appropriate in terms of timing and practicality.

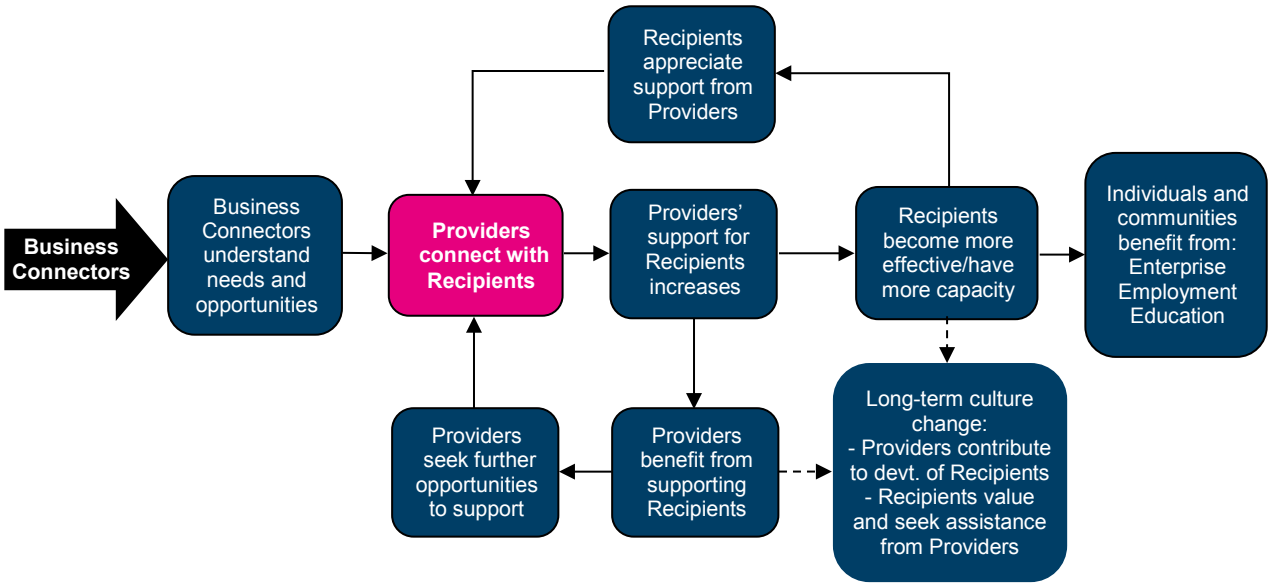
When developing a theory of change, you should put measurement aside. Begin by describing in qualitative terms what the organisation or programme wants to achieve, and for whom. This becomes the rock upon which your measurement framework is built. A theory of change shows the causal links between a charity's activities and its goals: it forces you to take a clear, simple view, crystallising your work into as few steps as possible to capture the key aspects of what you do and what you want to achieve. There is no set format to a theory of change, but we tend to describe it using a combination of diagrams and text.

Below we outline the steps in the process to developing a theory of change.

1. **Identify a realistic and definite goal.** Your goal is the end point of the theory of change—the ultimate aim of the charity or programme. It should be as clear and realistic as possible and identify who will benefit. If it is not clear, the theory of change will tend to expand into every possible activity and outcome that could happen, which it is not helpful for thinking about an appropriate strategy. If the goal is not realistic, it will be impossible to build a causal model of how to achieve it. At the same time, the ultimate aim should be ambitious enough to stretch the organisation.
2. **Work backwards from the goal to map the intermediate outcomes.** A charity typically achieves its final goal by working towards intermediate outcomes. For example, a charity may try to minimise the harm of cancer by sponsoring research, disseminating good treatment practices, and being an advocate for patients in policy decisions. These activities result in intermediate outcomes, as they are steps towards the final goal. Whether an outcome is intermediate or final depends on the scope and ambition of the charity. One way to identify intermediate outcomes is to work backwards from the final goal through each step, asking *'What has to happen in order for this to be achieved?'* This process will also ensure the focus is on what has to be done to achieve the goal, rather than on current activities. Box 1 explains the different types of outcomes.
3. **Establish the links between outcomes, and their order, by working out causes and effects.** It is important to go through the links in detail, questioning whether one outcome really leads to the next and why. The logic that you build up in this step becomes part of your impact story, so you need to make sure the links between the causes and effects are plausible and logical.
4. **Work out which activities lead to which outcomes.** You need to link the activities you undertake to your intermediate and final outcomes. These links should be logical and evidence-based. Leaps of faith should be avoided and suggest you may be missing some intermediate outcomes. What you are looking for is general agreement (though not necessarily complete consensus) on how the organisation achieves change.
5. **Identify what else is needed for the intervention to work.** A good way to think about this is to work out what would derail the intervention and what would accelerate change. This can reveal important enabling factors and barriers, such as whether stakeholders have to be on board, what policies would help, which partnerships you may need to foster, changes in the economic climate, and so on.

At NPC we applied this method to our work with Business in the Community, helping the charity to articulate how its Business Connectors programme leads to certain outcomes in the areas of enterprise, employment and education. The programme aims to change the culture of how business and communities work together to tackle social issues in their local neighbourhood. Business Connectors, who are talented individuals seconded from business, link up organisations—typically businesses and charities—where one organisation provides advice or resources which the other organisation receives. Figure 3 shows the high-level theory of change we created for the programme. Below this top-tier theory of change, we developed an individual theory of change for each of the seven roles that Business Connectors play.

Figure 3: Example of a theory of change



Box 1: Types of outcomes

When measuring change in people—such as skills acquisition, improvements in soft skills or changes in attitudes or circumstances—you generally have a choice of two types of outcomes:

Subjective, intrinsic, or soft outcomes are those that are valued by, and primarily relate to, individuals, such as self-esteem and confidence. These are changes relating to perceptions, feelings, attitudes or interpersonal skills.

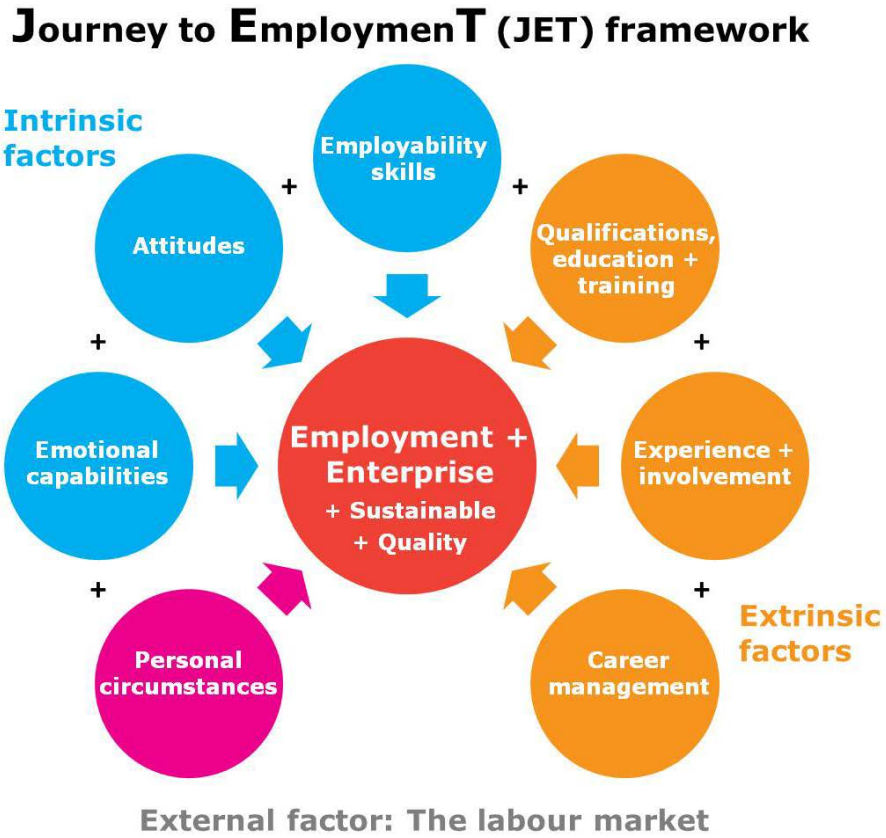
Observable, extrinsic or hard outcomes are those that are tangible, objective and can be more easily observed. These include educational achievement, literacy and numeracy skills, getting a job or reduced reoffending.

Similarly, when measuring changes in organisational practice or policy, you have a choice of:

- Subjective, intrinsic, or soft outcomes, such as views of policymakers and practitioners on your influence.
- Observable, extrinsic, or hard outcomes, such as counts of changes in legislation or observable changes in organisational practice.

The hard/soft and intrinsic/extrinsic distinction is useful, but there are grey areas. For example, psychometric tests aim to bring a level of objectivity to measuring how people think and feel. This can create some ambiguity and confusion in how these terms are used.

Both soft and hard outcomes can be intermediate outcomes or the final goal itself. For example, the Journey to EmploymenT (JET) framework below includes both hard and soft outcomes.[6] Soft (intrinsic) outcomes like attitudes and emotional capabilities are shown in blue, and hard (extrinsic) outcomes like qualifications and experience are shown in orange. The final goal of sustainable and quality employment contains both a hard component (being in a job) and a soft component (the quality of that job).



Pitfalls to avoid

Developing a theory of change may seem straightforward in theory, but it can be tricky in practice. Through our work with all kinds of organisations, we have identified some common pitfalls to avoid:

1. **Do not include elements that are neither activities nor outcomes**. For example, 'counselors are friendly' is not an outcome but a characteristic of staff. However, 'counselors are trusted by young people' is an outcome if it is a precondition for you to achieve your final goal and counselors have to work to build this trust. Features of your charity, the programmes you run, or the context in which you work are important, but these are not in themselves outcomes or activities. Your theory of change should focus on the changes that you expect to occur as a result of what you do.
2. **Do not overcomplicate things**. A theory of change should be, to quote Einstein, as simple as possible, but no simpler. Phrase outcomes in plain English. Focus on the outcomes and the causal links that are critical to the success of your charity.
3. **Do not over-claim outcomes**. Charities are ambitious about the change they want to see in the world and your theory of change should reflect that. But you also want to keep it simple and the outcomes you include proportionate to your resources and what you want to achieve. 'World peace' is an admirable aim, but not a realistic outcome in your theory of change.
4. **Always follow a logical flow**. Does your theory of change make sense? If one outcome does not logically lead on from another, you may need to include more steps in the chain, or describe the change differently.
5. **Do not include too many outcomes in the same step**. The best way to plot causal links between outcomes is for each outcome to capture one change; otherwise, your theory of change risks showing merely that everything causes everything else, and you will not be any the wiser as to what causes positive change. For example, it is probably best to split the following into two outcomes: 'Young people develop new skills and improve their self-confidence'. Combining outcomes may, however, be appropriate for a high-level theory of change.
6. **Do not include the same or similar outcomes more than once**. For example, 'service user makes better decisions' leads to 'improved confidence' which leads to 'better decisions'. Outcomes can be interdependent and causal links are not always linear, but you should aim to plot change in a simple and straightforward manner. If your model contains similar outcomes in more than one place, see if you can combine them into a single outcome or, for example, try revisiting the causal links and including feedback loops.
7. **Ensure your outcomes are specific enough**. 'Children have better outcomes' is too vague to be meaningful. Remember that you are working towards a measurement framework. To know if an outcome is happening, it must be well-defined. Ambiguous wording should be replaced with specific terms.

Final checklist

Sense check your theory of change by asking yourself if it is:[7]

- **Meaningful:** Does it describe the programme, project or organisation accurately in ways that staff, trustees, volunteers and stakeholders agree with?
- **Well-defined:** Is a clear audience, client or user group articulated? Is it clear what you do?
- **Comprehensible:** Does it enable you to give someone the "two-minute story" of the service? Would a member of the public understand the theory?
- **Doable:** Are the services and activities likely to contribute to the desired outcomes and impact?
- **Plausible:** Is it realistic? Does it take into account your organisation's capacity? It should be something that the programme, project or organisation could really do, not just wish it could.

- **Credible:** Are people outside your organisation likely to believe it? Is the secondary evidence you include credible with your stakeholders?
- **Testable:** Can you test the theory through a series of testable hypotheses? All elements should theoretically be assessed using research and observations of what happens (even though you may not have the resources to assess this yourself).

Case study: Shaw Trust

NPC helped Shaw Trust (formerly, the Careers Development Group) to develop a framework to measure and evidence its impact. Shaw Trust wanted to show the broader social impact of its work, beyond meeting its contractual requirements for getting people into sustained work as a prime provider of the Department for Work and Pensions' Work Programme.

Using our approach, we worked with Shaw Trust to develop a framework and relevant tools to show the difference it makes on jobseekers and their families, Shaw Trust staff and volunteers, other charities and the general public.

This involved two main stages. We conducted an initial scoping phase to understand how Shaw Trust perceives its social impact and to prioritise core outcomes to be measured. We interviewed six senior Shaw Trust staff, and held a workshop with a representative group of operational staff. We also conducted a review of Shaw Trust's annual reports, its output and outcome data, existing evaluation work, and of good practice relating to impact measurement in the employment sector.

This allowed us, together with Shaw Trust, to map the charity's theory of change—creating a clear conceptual model, which included the direct value of getting people into employment, as well as indirect beneficiary outcomes such as improvements in quality of life and well-being.

In the second phase, we worked in consultation with Shaw Trust to develop ways to measure the core outcomes from the theory of change. The tools we recommended included existing measures as well as bespoke ones developed specifically for Shaw Trust. We also created a reporting template to help Shaw Trust use the data collected to communicate their social impact.

Following the workshop and consultation, we finalised a social outcomes framework featuring outcomes, indicators and ways of measuring them. We also provided guidance on how to use the framework and a summary of lessons learnt from the process.

While Shaw Trust has not yet fully implemented the framework, the process was very useful to them as Special Projects Advisor, Nick Carey, explains: *'The work NPC performed on developing a theory of change and measurement tools for Shaw Trust was hugely useful. Not only did it assist us in describing and measuring our impact in a systematic and objective way, it also helped reinforce the link between our mission and our contractual delivery activities. The importance of having an impact-focused organisation cannot be overestimated, particularly in the current environment, and working with NPC was a most educative and enjoyable experience.'*

STEP TWO: PRIORITISE WHAT YOU MEASURE

Building on what already exists

With your theory of change in place, and with that all-important buy-in from across your organisation, you are ready to move onto the next stage: selecting the most important outcomes to measure.

You may not need to measure all of your outcomes in order to understand the impact you are having. A charity's impact is likely to be diffuse and charities have an effect on different people in different ways over different time frames. A parenting support programme may help parents directly, but also the children, grandparents and ever-widening circles of beneficiaries over the short, medium, and long-term. Trying to capture all these changes is complicated and may not be a good use of scarce resources. Charities therefore need to think hard and focus on measuring the essential outcomes.

We have seen many organisations be overambitious about what they plan to measure, resulting in an unwieldy system that produces large amounts of data of variable quality, much of which goes unused. We advise not overcomplicating it: you want to get to that Goldilocks point of collecting the right amount of quality data. You might require some trial and error to get there, but taking good decisions on what is important to measure is critical. The framework we developed with Shaw Trust (see case study on page 16) helped the organisation to clarify the relative importance of its potential impact on a range of stakeholder groups, and then focus its data collection efforts.

It can also be tempting to collect and report data that is convenient rather than important, and to let this drive what you measure. This is a sure route to collecting data that is not useful, leading to data quality that suffers over time as management attention to the data drifts.

While you will naturally focus on positive impacts, do not ignore the possible positive or negative unintended consequences of what you do. Try to identify where you may have a negative impact, and consider tracking that. Not only will this help you to improve, but if you do not understand the possible negative impacts you may have, you are vulnerable to criticism from someone who does.

Deciding what to measure

To decide on which outcomes to measure, ask yourself two questions:

1. Has anyone already proved the causal link between outcomes in your theory of change?

If so, you can use desk research to gather this evidence rather than collect new data. For example, an organisation that aims to improve children's academic achievements may have these children growing up with better employment outcomes in the long term as part of its theory of change. There is research to help back up causal links between higher academic achievement and employment outcomes, so the organisation could focus on tracking academic success rather than try to follow up the children they work with into adulthood, whilst acknowledging that children would ideally be followed up.

2. Is it really important for you to have data on this outcome?

You need to prioritise the most important outcomes. These will inevitably need to reflect the outcomes that your stakeholders, especially funders, see as important.[*] Nevertheless, they should also be outcomes that:

- you **directly** influence (rather than **indirectly** support);
- are **important** or **material** to your mission;
- are not too **costly** to measure; and
- will produce **credible** data.

This process will help you prioritise a set of outcomes to measure. These outcomes form the basis of your measurement framework.

What should I include in a measurement framework?

At a minimum, your measurement framework should include:

- a list of the outcomes that you want to measure;
- indicators or proxies of those outcomes (ie, something you can measure that informs you if that outcome has been achieved); and
- sources of data of the indicators (ie, where the data will come from).

The framework could include other relevant information, such as the measurement tool you will use, the method for analysing the data, the frequency of data collection, and known limitations of the data.

[*] Sometimes charities are asked by funders to report data on outcomes that are not relevant to the charity's mission or objectives. Different funders will also have different requirements regarding the outcomes they would like data on. NPC's *Turning the tables* report looks at reporting requirements for charities, which can be overly burdensome and excessive, and calls on charities and funders to work together to streamline reporting. A theory of change helps charities to negotiate with funders on the priority outcomes to collect data on.

Case study: Business Connectors

Business in the Community (BITC) is a business-led charity committed to building resilient communities, diverse workplaces and a more sustainable future. BITC asked NPC to help develop a framework to measure the social impact of its new Business Connector programme.

Business Connectors are talented individuals seconded from business, trained by BITC and placed in local communities of greatest need. They encourage collaborations among a wide range of organisations and spend time with key people to better understand the issues facing their community.

Overall, the Business Connectors programme aims to change the culture of how business and communities work together to tackle social issues in their local neighbourhood. The target is to create a network of Business Connectors over five years, supporting over 200 local communities across England.

Capturing the impact of the Business Connectors programme is challenging. The goal is, as the name implies, to create and/or support inter-organisational relationships that address local issues. Of course, these differ by area and the ultimate impact of the programme rests on the extent to which these relationships bring benefits to these communities.

We took BITC through our approach, firstly conducting a literature review on the value of brokerage and of business engagement in communities. This informed the theory of change for the programme, which was developed in collaboration with BITC, the Big Lottery Fund, charities and businesses. For a programme that affects many organisations in different ways, it was important to measure a few priority outcomes well, and so the theory of change was absolutely central to deciding what these were. We then worked with BITC to identify indicators for these outcomes, which included the number of different types of connections made and their usefulness, as reported by local charities and other stakeholders.

After this, we developed a range of tools for data collection, which involved modifying the Business Connectors' management information system, creating surveys for the organisations that provide and receive support, and giving guidance on selecting and developing case studies that illustrate the broader social impacts. We also trained BITC's staff and Business Connectors in using these tools.

BITC say this approach has helped them to articulate the aims of the programme clearly, improve their training of new Business Connectors, increase the effectiveness of the programme and demonstrate to funders and stakeholders the difference it makes.

Lisa Cunningham, the Programme Director of Business Connectors, reports that developing and implementing the measurement framework has been '*hugely beneficial to the programme'*, adding that '*we have benefitted from NPC's expertise and experience.'* Andrew Morris, Deputy Director England of BIG Lottery Fund, a major funder of the programme, said: '*These tools not only provide high quality and real-time data to BITC to help them manage their programme, but also allow Big Lottery as the funder to see and measure the impact of our funding in a clear, consistent and compelling way'.*

STEP THREE: CHOOSE YOUR LEVEL OF EVIDENCE

Ways to show you make a difference

A theory of change helps you choose what to measure, but before you actually begin measuring you must decide how rigorous and credible your evidence needs to be.

This is not straightforward. Different stakeholders will have different views about what counts as strong or credible evidence. Frontline staff see the benefits of a service or programme in their day-to-day work and may need little convincing that what they do makes a difference; but government researchers and seasoned academics may be persuaded only by very rigorous evaluations. Finding the right balance will depend on a number of factors: while you may want to create strong evidence of your impact, you will be limited by resources, and by practical and methodological constraints. This guidance will help steer you in the right direction and avoid either under-investing or over-investing in measuring your impact.

There are four main ways[*] to make a credible case that what you do really makes a difference (see Appendix 3 for more detail):[†]

1. **Statistical** approaches look for patterns in quantitative data to see if the 'effect' (ie, the expected outcomes) frequently follows the 'cause' (ie, the service). Such approaches include 'before' and 'after' comparisons, correlation, regression analysis, and other statistical models.

   Example: *Place2Be compares the mental health of children before they receive its therapeutic services to their mental health afterwards. Large numbers of such comparisons show improvements in the before and after comparison. As there appear to be no other reasons for this consistent pattern, Place2Be conclude the most likely cause of these improvements is the effect of their counselling sessions.*

2. **Experimental** approaches compare differences in outcomes between people who receive a service (the 'intervention group') and people who do not (the 'control group'). The most rigorous way to select a control group is to randomly choose who receives the service and who does not in a randomised control trial (RCT). Any differences in the outcomes of the control group can be attributed to the service or to chance using statistics. This approach helps distinguish a programme's impact from other factors that might affect the outcomes. The most common concerns with RCTs are the cost and ethical concerns about denying people a service.

   Example: *In partnership with the local authority and with two universities, the Brandon Centre ran the first UK randomised controlled trial of Multisystemic Therapy (MST) to test whether MST is more effective than "business as usual" in preventing young people aged between 13 and 17 from reoffending. The trial ran from*

[*] Befani, B., Davies, R., Forss, K., Mayne, J., Stame, N., Stern, E. (2012) *Broadening the Range of Designs and Methods for Impact Evaluation*. Department for International Development, Working Paper 38. The article includes participatory approaches as a fifth approach. Participatory approaches go beyond beneficiaries participating in assessing impact, which is a well-established practice, but concern who decides what impact means and how it should be assessed. Such approaches are not widely used to explain to external parties what difference a programme makes and so we have not elaborated on it here.

[†] Evaluators distinguish between attribution and contribution. Attribution addresses whether an intervention or service *causes* an impact, and contribution addresses whether an intervention or service *helps cause* an impact through combining with other causes. It makes more sense to talk of attribution when there is a clear link between what a charity does and the impact, and to talk of contribution when the link is indirect. While a useful distinction, in our experience it is not widely used by funders and charities in assessing their impact.

2004 to 2010 and involved 108 families. Participants were randomly assigned to the two groups. The findings showed statistically significant lower levels of the risk of reoffending, and in the number of offending behaviours, for those who received MST compared to the control group who did not.

3. **Case-based** approaches compare cases (eg, individuals, groups of people, or places) *within* an intervention or programme, or *across* interventions or programmes, and use the similarities and differences between them to draw conclusions about causes of the effects and the impact of programmes. This approach recognises there is rarely a single cause to any social outcome and that it is difficult to unpick the influences of each of them.

   Example: *StreetChance is a cricket programme run in deprived areas in different cities across the UK. The aim is to foster and support pro-social and healthy attitudes among the participants while also promoting cricket. NPC's recent analysis found that participants in areas where the local police are involved in the programme are more likely to have pro-social views than those in areas where the police have not been involved. There could be several reasons for this: it could be coincidental, or police involvement may deter young people who have more anti-social views or help increase understanding and trust, as the programme hopes to do. Careful analysis of data collected over time will show which is the most plausible.*

4. **Theory-based** approaches describe in detail how a service or programme influences different people at different times and places using observations by staff, evaluators, and other stakeholders, as well as what beneficiaries say, rather than by analysing lots of cases or using a control or comparison group.

   Example: *Guy's & St Thomas' Charity funded a theory-based evaluation of a multi-agency domestic violence service that was based in maternity and genitourinary services at Guys and St Thomas' NHS Foundation Trust. The evaluation, undertaken by King's College London, examined the key assumptions about how and why the intervention would work and tested them using several sources of data, including pre- and post-training questionnaires, focus groups/interviews with health professionals, audits of patient records, and quantitative and qualitative interviews.*[*]

These different types are not mutually exclusive; they can be combined to bolster your claims that that a service or programme makes a difference. Appendix 3 provides common methods used for all four.

Note that we have not included econometrics (the use of quantitative data to test and predict impacts based on economic theory) or economic evaluation (a comparison of the costs and consequences of services or programmes, typically using monetary values) in this guidance. These economic models refer to specific applications of the Statistical and Experimental approaches. Both rely on robust quantitative data and are therefore not always appropriate. We plan to publish separate guidance on when charities and social enterprises should and should not use economic evaluation as a way of assessing and describing their impact in due course.

[*] More information may be found at: http://www.caada.org.uk/policy/bacchus-et-al-2007-full-report-mozaic.pdf

Which approach is 'best' to demonstrate your impact?

The evidence each approach produces is not equally credible—or perceived to be equally credible—when it comes to backing up claims about the impact you make. Social scientists have created scales or levels of evidence that reflect this.[8] Two such scales used in the UK charity section are described in Appendix 2. For both:

- The experimental approach is regarded as the best at avoiding bias, including the optimism bias in interpreting and presenting data that is tempting for charities as they compete for funds. Hence, randomised control trials are often held up as the "gold standard" of evidence.
- Statistical approaches are regarded as the next best, although how good they are depends on the quality of the data and the type of statistical analysis you apply.
- Theory-based approaches are viewed as the weakest if they only use qualitative data and do not include elements of other approaches.

However, not all social scientists and evaluators agree that such scales are useful or that randomised control trials always provide the best evidence.

The scales work best for clearly defined programmes where there are many cases (individuals, groups of people, places), and do not work as well for programmes with one or a few cases in complex and changing environments, such as community development initiatives. Critics point out that randomised control trials do not always address the right questions. Furthermore, in many situations control groups cannot easily be applied[*] and it is hard to allocate people randomly to the intervention or control group when, for example, you are targeting small groups, the hardest to reach, or people in a specific geographic area. Finally, like any method, trials can also be subject to optimism bias.

Ultimately, there is no hard and fast rule as to which approach is best for any given programme or service. We provide guidance to help charities work towards a decision. Our main recommendation is to take the challenge seriously: do not simply take the 'easiest' option—it may not give you results that will convince your stakeholders.

Each approach has strengths and weaknesses, outlined in Figure 4.

[*] Practical options for setting up control groups are given in Appendix 4.

Figure 4: Potential strengths and weaknesses of different ways to show causality[9]

Approach	Potential strengths	Potential weaknesses
Statistical	• Simple to collect and analyse data • Can identify patterns of apparent cause and effects for further investigation	• Requires many cases (eg, cannot be used to assess the effect of a charity on system or policy change) • Most credible in a simple or stable environment with a simple and plausible cause-effect relationship • Does not explain why or how change happens
Experimental	• Avoids bias • Can produce strong support showing whether the service/programme being assessed makes a difference (strong internal validity)	• May not be appropriate where context and underlying conditions are important • May not be able to generalise from the results (weak external validity) • Does not explain why and how change happens
Case-based approaches	• Can uncover typologies/sets of causes for effects	• Does not work in very complex environments that have too many possible sets of causes and conditions to consider
Theory-based	• Can provide good evidence of why change does, or does not, happen, and for whom • Takes context, underlying conditions, and complexity into account	• May not measure size of any effect • Susceptible to bias/self-serving interpretations of data

How much should you spend measuring your impact?

Our 2012 survey on impact measurement practices among charities in the UK suggests that on average (though with a degree of variation) charities spend 3% of their total budget on monitoring and evaluation—considerably lower than the 10% that is sometimes recommended.[10] But the question above is the proverbial 'how long is a piece of string?' Instead, ask yourself:

- How much evidence do you require to make decisions in your life?
- How much time should you invest in figuring out which is the right decision?

Clearly, there is no single answer to either. The context and scope of the decisions is all important. Buying a house is quite different from deciding what to cook for dinner. Similarly, choosing how much to spend on measuring or evaluating your impact depends on the context and scope of what your charity does and should be relative to the potential benefits, the external risks, your 'risk preference', the short- and long-term consequences, the availability of information and so on. If not sure where to begin, ask yourself *'How valuable to you and your beneficiaries is information that tells you how to improve your impact?'*

Selecting an approach and methods

Your choice of approach(es) will depend on a combination of what is desired or needed and what is practical. Here are some general guidelines to follow:

- **Make your evidence as rigorous as possible by creating a credible counterfactual:** Any attempt to create a picture of the world as if your charity or programme did not exist is to use a counterfactual. For example, a 'before and after' comparison, where the 'before' situation is unlikely to change except for the intervention, qualifies as a counterfactual approach. But a strong counterfactual approach compares the outcomes of people who participate in a programme or service (the intervention group) with similar people who do not (a control or comparison group). Depending on how the control or comparison group is selected, statistics allow researchers or analysts to estimate the likelihood that the programme or service caused any difference between the outcomes of the two groups.

  Other things being equal, the experimental approach provides the strongest evidence for people who are unfamiliar with your service because it creates a clear counterfactual that distinguishes the impact you make from other factors.[*] But randomly selecting who receives your service may create practical problems and raise concerns about being unfair, or worse, harmful towards those who are deprived of it.

  Even when it is possible to create a control group, however, it is often difficult and expensive to do so and many charities find it hard to justify this cost to funders. Moreover, it is not necessary to run an experiment to examine different explanations. Whatever approach you take, you should pay close attention to exactly who you help, how and why, using the principles of theory-based approaches.[†]
- **Consider and rule out any plausible alternative explanations of the impact you think you have:**[11] Naturally we all like to think we make a difference: but the claims you make to others—some of whom might be sceptical—will be stronger if you can refer to and rule out other explanations. In the classic example, the former Mayor of New York, Rudy Giuliani, attributed the drop in crime across the city in the 1990s to his "broken windows" approach; but criminologists point out that crime fell sharply nationwide, particularly in big cities, and that there were other social, economic and demographic factors involved. It is impossible to fully rule out other explanations, but the more you try to the stronger your case will be.
- **Choose an approach to help you achieve your mission, not just to benefit your organisation:** The biggest barrier that charities face in producing credible evidence of their impact is the pressure to produce results that show the organisation in a good light, rather than contribute to an understanding of what does and does not work. While this is a natural and possibly unavoidable consequence of the competition for funding, your first priority is to help your beneficiaries. You can do this by creating and publishing evidence of what you do.
- **Take seriously the possibility that you might be doing harm:** The greater the risk that you might cause harm, the more robust your approach should be. There are many examples of medical practices that were theoretically sound and appeared to benefit patients in practice, but when put to robust methodological tests were actually found to cause harm or make no difference.
- **Tailor your evidence to the needs of your stakeholders:** For your evidence to be persuasive, it needs to withstand criticism. You may want to get it reviewed by someone outside your organisation who can judge it independently. If you are trying to persuade government departments to make a big investment or a change in policy, you will need very credible evidence, most likely using either the experimental or statistical

[*] It is technically better to talk of *estimating* impact rather than *measuring* impact, as a charity, programme or service's impact is the difference between the outcomes that occurred compared to what would have happened otherwise. We can only estimate what would have happened otherwise, so impact is the difference between actual and estimated. However, here we follow the convention of using measuring impact and estimating impact interchangeably.

[†] Denying people your service might be unfair if they are offered nothing else instead and you have strong evidence that your service does help, but the concern might also just reflects optimism bias that your service is better than what would happen otherwise. Beware—this optimism bias can and has hurt people.

approach. If you want to explain to funders and internal audiences how to improve on disappointing results, you should focus on exactly how your service affects people using the theory-based approach.

- **Take the approach that is *accepted* among your peers and within your sector as a minimum and try to do better:** This is a useful check on the minimum level of rigour that you need. For example, medical research charities find higher levels of evidence—such as randomised control trials—are typical, while in the education sector, analysing correlations between participating in a programme and educational attainment (ie, the statistical approach) is more common. The accepted norm should be your starting point. To stand out you will need to do better.
- **Build on evidence that has already been developed:** Before you start collecting data, make sure you are familiar with the relevant academic literature. If existing evidence for programmes similar to yours is already strong, you need only collect data that shows your programme is likely to replicate those results. But if your service is innovative and there is little in the way of existing evidence, you will need to increase the rigour of your evidence over time to persuade yourselves and others that your service works. Accessing and understanding academic literature is not simple, so you may need support to do this.
- **Avoid costly experimental or quasi-experimental methods if your service is not yet stable and consistent:** If your service is new and you are reviewing it as you learn, you will need short-term feedback loops to help you improve: focus on getting these short-term measures right. A robust, costly, long-term impact evaluation will be inconclusive for services that are changing.
- **Consider the resources—the time, budget and skills—you have available**: Improving the strength of your evidence will typically—but not always—require putting time and money towards planning, collecting data, and analysing the results. It is necessary to have some understanding in your organisation of what is more and less credible evidence, but hiring an external consultant can help you design an appropriate monitoring and evaluation programme to ensure your analysis is rigorous. This comes at a price but may be the most cost-effective approach.[*] Hiring someone to design an appropriate approach is different from hiring an independent evaluator of your services.
- **Choose validated measurement tools rather than tools you create yourself**: While cost and practicality are important in choosing tools, you will need to consider the rigour of the tools. You should prefer tools that are standardised and validated: standardised tools measure the same thing for different people and organisations across different contexts, and so results can be compared, and validated tools produce results you can trust. NPC's well-being measure is a standardised and validated tool. It has been tested across multiple schools and validated externally. Not all existing tools are standardised and/or validated as this can be time-consuming and difficult. The benefit of developing your own tools is that you can customise them to your needs. The downside is that they may not be viewed by researchers and academics and some funders as rigorous and validated tools.[†]
- **Consider tracking medium and long-term changes to improve the evidence of your impact:** The difficult part of tracking long-term impact is typically keeping in touch with users. This should not put you off.[‡] One reason why charities find this hard is because they do not realise how useful it can be and so do not prepare for it. For example, you can ask users while you are working with them if they are happy to be contacted once the programme has ended. Also consider collecting contact information of family and/or friends to help track people once you have lost contact with them. An alternative to following-up with users directly is getting hold of secondary longitudinal data—long-term data that someone else (usually researchers or government) collects about your beneficiaries. Unfortunately, there is not a huge amount of this available and it may be hard to get hold of.[§]

[*] For instance, an externally commissioned RCT could cost anything between £20,000 (for a simple trial) up to hundreds of thousands of pounds.

[†] For further guidance on use of tools see: http://www.clinks.org/sites/default/files/UsingOffShelfToolstoMeasureChange.pdf.

[‡] For guidance on this see: http://www.clinks.org/sites/default/files/AchievingUserParticipationResearch.pdf.

[§] Start by looking at the UK Data Service: http://ukdataservice.ac.uk/.

- **Look for different ways to strengthen your theory of change:** The strength of the evidence of your impact depends on the strength of the causal links in your theory of change. The more evidence you have of those links, the more confident and persuasive you can be about attributing any changes to your service. So if you cannot create a control group or run a trial, you can strengthen your evidence by: 1) using other people's research; 2) collecting and using good qualitative data to support any quantitative measure of change; and 3) using more than one source of data to support each causal link in your theory of change. The latter is known as triangulating data and helps avoid any bias, or perception of bias, from using only one data source.

Example

Suppose your theory of change is that by increasing the opportunity to engage in sports at school you help lower rates of obesity in schoolchildren and reduce health problems in later years. This seems logical, but you face a problem tracking future changes in health and obesity. You might be able to get access to NHS data on individuals you have worked with (subject to confidentiality), but you should also look closely at your theory of change. Suppose there is evidence that, in general, participation in sports at a young age is associated with fitness at an older age, but also that sports at school tend to *exacerbate* fitness and health problems as PE teachers tend to focus on the most sporty students and the less sporty ones get discouraged. You now have a different, but easier, case to make. You can use the existing evidence of the importance of sports at a young age to support your service, but you need to demonstrate that your service leads to increased enthusiasm for sports and an uptake in sports by children who are least keen and normally unlikely to participate. By doing this you will have fitted your theory of change into existing research to show where your small change can make a big difference.

Case study: Women's Counselling and Therapy Service (WCTS)

Based in Leeds, Women's Counselling and Therapy Service (WCTS) provides counselling and psychotherapy services for vulnerable or disadvantaged women. In 2012, it began a new project called MumsTalk, funded by The Tudor Trust. MumsTalk offers counselling to mothers and female carers dealing with difficulties that affect their ability to nurture themselves or their children, with a particular focus on mothers whose children also receive therapeutic support.

The charity asked NPC to help it develop a measurement framework for MumsTalk, so it could offer both funders and clients a coherent picture of the benefits of the service.

Creating a theory of change was central to the whole process. We held a workshop to pin down what the service aims to achieve for both mothers and children, and through a collaborative and iterative approach involving representatives across the organisation, a final version was agreed. This provided clarity around the goals of the project, allowing us to establish the flow of outcomes and identify which to prioritise for measurement. Stella Maden, who runs MumsTalk, found the theory of change exercise particularly beneficial in helping to distinguish between outcomes for the mother and outcomes for the children (through working with the mother), and the charity has now adopted the approach for other projects.

NPC used the theory of change to guide research into appropriate data collection tools. WCTS was already using the CORE outcome measure (a measure of psychological distress) across all its services, and so we focused on looking for additional tools that could be used alongside. As a result, two new measures are now being used for a before and after comparison (the **statistical** approach), and the range of evidence captured has now been broadened to include the parent-child interaction. NPC also helped WCTS with changes to its feedback form, including questions on the clients' perception of the difference that the service had made to their difficulties, and on other events in their lives which might have had an impact on their recovery.

STEP FOUR: SELECT YOUR SOURCES AND TOOLS

What type of data do you need?

Once you have identified the level of evidence you require, you need to think about how to gather your data. You may be able to use existing tools or data produced by academics or other organisations, or you may need to develop your own. You should look at what already exists and assess its suitability—it is likely others will be measuring the same outcomes as you, and you do not need to reinvent the wheel. You can only assess this suitability once you have developed your theory of change and are clear on the change you seek, how you plan to make it, and what you want to measure. You may also be able to use existing evidence relevant to a causal link in your theory of change, to avoid having to measure that change yourself, or to support measures you put in place.

The measurement frameworks NPC helps to develop for organisations typically combine tools that have been developed by others with tools customised to fill in the gaps. Ultimately, it is important to use measurement tools that are fit-for-purpose and capture the change you want to bring about. Do not be tempted to prioritise data that is easy to collect but does not tell you much about your impact—ultimately this will be a waste of resources. And remember that measures of impact are proxies for changes in the real world—they are not perfect. You need to be constantly vigilant that the measures you use illuminate rather than obscure the changes you make in reality.

The outcomes you prioritise and your decision about how rigorous your evidence needs to be will determine the type of data you should collect and how you should go about it. There are three key questions to answer:

- **Who** do you need to collect data from?
- **What** type of data should you collect?
- **When** should you collect this data?

Who to collect data from

Who you collect data from should reflect the groups your organisation interacts with—the beneficiaries of a service, staff members, practitioners using your resources, or policymakers you are trying to influence. But it may also be appropriate to consider groups with whom you have indirect contact, such as beneficiaries of an organisation that you directly support. Using a control group will involve carrying out research with groups *not* using the service—either collecting data from them or another organisation, such as a government department. You may also want to contact people who choose not to use your service to understand why they do not. In addition, the views of external stakeholders, such as funders or peers, may provide insight into the impact of your activities.

Tracking a sample of people rather than everyone you work with cuts costs without sacrificing the quality of the evidence. This is a must if you work with a large number of people. If you sample properly, you can estimate what happens to people who are not in your sample. The main methods for selecting a sample are: random, quota, purposive, and convenience. Each has its pros and cons, and the best choice depends on what you are trying to find out. Appendix 5 includes more detail on these approaches, and how to select a sample. More information is covered on NPC's/Clinks' *Improving Your Evidence* website.[12]

What type of data to collect

Data is most commonly categorised as quantitative or qualitative. Quantitative data helps you to determine whether change has taken place, while qualitative data helps you to understand why and how. Quantitative data is aggregated evidence of the extent of the impact you are having. Qualitative data, including case studies, is more effective in communicating the detail of what has or can be achieved and how. The data types are complementary, and NPC promotes the collection of both where possible, bearing in mind the mantra: '*No numbers without stories; no stories without numbers*.'

The table below illustrates the differences between quantitative and qualitative data.

Figure 5: Qualitative and quantitative data

Type of data	Type of findings	Source of data	Example finding
Quantitative data	Statistical estimates for the prevalence of views, attitudes and experiences	Questionnaires and administrative databases	*62% of our members found our newsletter useful.*
Qualitative data	Detailed understanding of how and why an intervention has led to change	In-depth interviews and focus groups	*One of the main reasons members found our newsletter useful was because it was very informative on the latest policy developments.*

Although this guidance focuses on measuring outcomes and impact, you will most likely use some data on your outputs to inform your theory of change. Outputs are the amounts of a particular service or activity that an organisation delivers. Some output data may not be directly relevant to your impact—for example, knowing how many counselling sessions a person has attended may be important when it comes to assessing the service, even though it tells you nothing about whether their well-being has improved. There is sometimes a degree of ambiguity about what counts as an output or an outcome. For example, class attendance may be an output for a particular education programme, but for another working with young people who are disengaged from school, attendance and participation may count as a success and may therefore be considered an intermediate outcome.

It is good practice to triangulate data. This involves comparing the findings from one type of data or source (such as beneficiaries of a service) with the findings from another (such as practitioners) to get a better understanding of impact. This might also involve triangulating quantitative data with qualitative data. For example, in assessing the impact of its music therapy programme on children with communication difficulties, the charity Coram compares the views of parents with the views of practitioners. Practitioners will notice subtle changes that are related to the therapy, but what they see is limited to what happens in the sessions. Parents may not be as aware of the specific changes the therapy is making, but they see their children in settings outside of the therapy.

When to collect data

A key decision to make is when and how often to collect data. Resources, funder requirements and the types of activities you carry out, as well as methodological and practical considerations, will all influence this decision. You need to think about:

- **Collecting baseline data**. Baseline data is information gathered before a programme begins to capture the situation prior to your intervention. It provides a point of comparison for assessing your programme's impact. Without good baseline data you cannot make a reliable comparison and you risk underplaying your impact. This principle applies to all methods of evaluation.

- **The frequency and duration of data collection**. Will you collect data on all your activities at all times, or carry out smaller, one-off evaluations at regular intervals? For example, a telephone helpline may survey all callers for one week every quarter to complement routine call volume data.
- **The timing of data collection in relation to the beneficiary's experience of the intervention**. A common pattern is to collect data before and after an intervention. This requires you to think about how long after the initial data collection you can expect to see a change in outcomes.
- **Whether and when to collect follow-up data**. Charities seeking to bring about long-term change should consider collecting data from beneficiaries months or years after the intervention. Without this kind of longitudinal data it is difficult to make claims about long-term impact. In some instances, it may be possible to get follow-up data from statutory agencies—for example, on employment or offending. In practice, it can be difficult to find out who holds what statutory data and gain permission to access it.

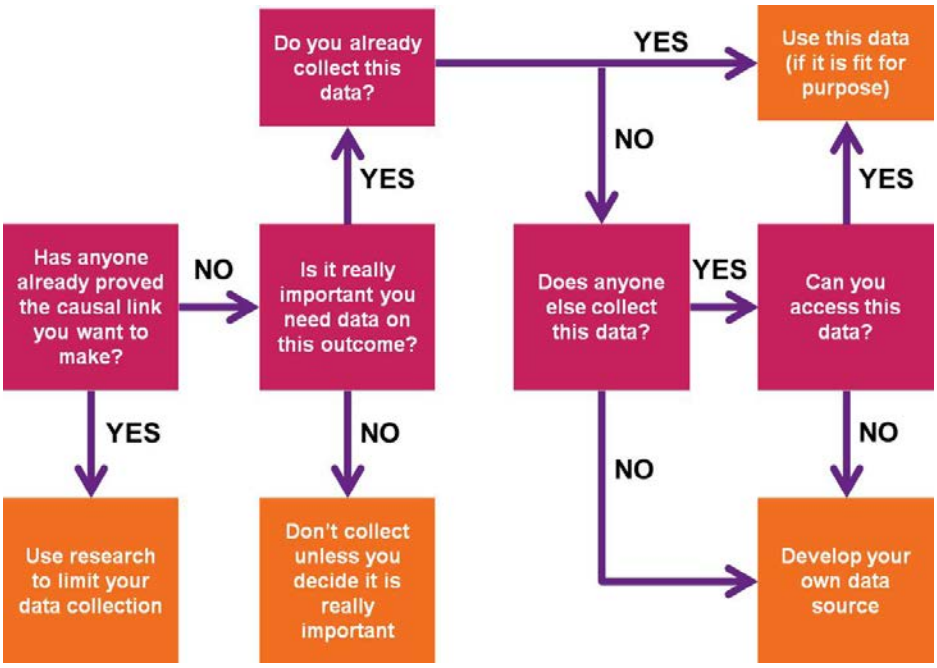
How will you get the data?

The final stage in developing your measurement framework is to identify how to get your data—the approach you take and the tools you use to collect it. First, you should think about what data already exists and how to use that where possible. Then you can identify new sources of data and find or develop appropriate tools to collect it.

Using what exists

Using existing evidence, data or tools—where appropriate—will make measuring your impact more effective and efficient. You will have already identified existing evidence that can help support your theory of change, and prioritised the outcomes you need to measure: now, by identifying useable existing data, you can limit the level of resources you will need to use to collect new data (see Figure 6).

Figure 6: Deciding how to get your data



When considering the usefulness of data that you or someone else has already collected, bear in mind two things:

1. The quality and robustness of any existing data needs to match the level of evidence you are aiming for and adequately capture the outcomes in your theory of change. You may have to modify your existing data, for example, by adapting an existing research design or set of questions—do not just accept what you already have.
2. There may be barriers to accessing data that others collect. This may be due to data protection restrictions, resource constraints, or bureaucratic hurdles. Even if the data you need is already being collected, there will be situations where it is easier to collect it again yourself.

Developing your own data source

Charities typically need most help when they have to identify a new data source and find an existing collection tool or develop a new one.

Sources of data

Data can come from a number of different sources. You need to be clear on the source of your data, as this can affect what inferences you can draw from it. The three most common are:

1. **Primary data**: information provided to you from people you directly work with and impact.
2. **Secondary data**: information collected by someone else. This includes data in academic literature; research undertaken by government, think tanks, or other research institutions; or data held by public sector agencies or other organisations on issues like employment or offending. It is important to know how these third parties collect their data so that you can interpret and use it correctly.
3. **Informant data**: where you cannot get information from the individuals you directly impact, you can consider obtaining data from an informed individual—someone well-placed to comment on any change in a beneficiary. Informant data often comes from parents or other relatives, teachers, health practitioners or other professionals working closely with an individual. Even when it is possible to obtain information directly from the person you work with, data from other sources can be useful to triangulate your results.

Choosing or developing a tool

For some data sources—for example, when you use secondary data—you will not have control over the tool that is used for collection. If you want to know the reconviction rates of ex-offenders that you work with, and have a public sector source for this (such as the Justice Data Lab, described in Appendix 6), the tool is already given—in this case, a reconviction decided by the courts and recorded in government databases. But for other outcomes and sources, you will need to choose which tools to use. Tools can be split broadly into two categories: quantitative tools and qualitative tools (see Figure 7). Appendix 6 describes different types of tools in more detail, and how to assess their suitability for your requirements.

Figure 7: Quantitative and qualitative methods and tools

Common quantitative methods and tools	Common qualitative methods and tools
• Surveys/questionnaires • Psychometric scales • Case work tools/records • Statutory data	• Interviews • Observation/ethnography • Focus groups • Anecdotes and feedback

Figure 8 presents a number of issues to consider when choosing the best tool for you.

Figure 8: Choosing tools

Area	What to consider when choosing tools	NPC's tips when choosing tools
Outcomes	• The mix of 'soft' or 'hard' outcomes • The fit between the outcomes you want measured and the tool • Both measuring change and understanding why change has happened	• Choose standardised and validated tools where you can • Try to get hold of existing administrative/ statutory data for hard outcomes • Use quantitative tools to measure change and qualitative tools to understand how change happens
Activities/ intervention	• Do you work with people over a long period of time? • Do you work with people one-to-one or in an intensive way? • Is your service innovative or established? • Are you trying to scale your approach?	• Consider how you might follow up with people in the long term, such as contacting a random sample or using statutory data • Resource-intensive tools may not be appropriate for light-touch interventions • If you are delivering a new activity or trying to scale you should invest in a robust approach to measuring your impact
Beneficiaries/ stakeholders	• How accessible are your users/beneficiaries/stakeholders? • How easy is it to get a representative sample? • How easy is it for respondents to take part in research?	• Consider secondary data sources for beneficiaries/stakeholders who are not accessible • Ensure the tools are tested for the population you work with—eg, older people, people with learning difficulties
Time and resources	• How to minimise the collection of new data (ie, using existing evidence and data collected by others) • Available resources to collect, use, and analyse data	• Identify a lead in your organisation to drive measurement of your impact • Recruit someone with relevant skills and experience as needed • Research tools that others in your sector use rather than reinventing the wheel • Make use of training and free-to-use tools and guidance • Consider sharing measurement with other organisations to reduce costs[13]
Need for rigour	• What is widely accepted in your sector? • What is your appetite for rigour • The availability of tested tools for outcomes you seek	• Look for shared measurement and common tools to increase rigour • Research availability of tools before developing your own
Funders	• What is your funder's appetite for evidence? • Are your priorities for measurement aligned with your funder's? • Are your different funders asking you for different information?	• Talk to your funders to understand their priorities • Talk to funders who ask for information that is not a priority for you • Work with funders to reduce duplication in reporting. Consider developing a shared measurement approach with charities and funders in your sector to streamline reporting.

Piloting your tools

Piloting is an important step before you use any new tool. Piloting simply means testing your tools to make sure they work in the way you want them to. It has a number of different purposes:

- Making sure that whoever provides the data (eg, respondents to a questionnaire) gives you what you need. Asking a sample of respondents to talk through their answers as they fill in a questionnaire is known as cognitive testing.
- Checking to see if there are any mistakes or glitches in the tool.
- Testing the time it takes to respond or carry out the measurement.

You should pilot your tools with a sample of people who represent the population you want to get information from. If you are surveying a group of young people, for example, you should pilot the survey with a small sample of young people. If you intend to use the tools with a small group (30-50), you should pilot with three or four people. For larger samples, try to pilot with at least ten people.

Case study: Send a Cow

Send a Cow works in seven countries in Africa, giving communities and families the hope and the means to secure their own futures from the land. The charity asked NPC to help it refresh its approach to measuring social impact, concerned that the conventional method of reporting impact through logframes and funder-determined 'hard' outcomes, such as income and agricultural productivity, did not properly reflect the notable social and personal changes that practitioners were seeing on the ground.

NPC and Send a Cow combined their experience and expertise in impact measurement and international development, respectively, to co-develop a framework that captures the changes the charity aims to bring about. First we ran a theory of change workshop with participants from the different countries in which it operates and from different roles across the organisation.

Through a process of comment and review, we managed to refine the theory of change until there was consensus. We were then able to research indicators of the changes its services hope to achieve and data collection tools, drawing on both the UK and international development sectors.

Send a Cow is taking a staged approach to implementing the framework, using data it already has while introducing new tools to collect data it still needs. The results to date can be seen in its impact report at www.sendacow.org.uk/our-impact.

Richie Alford, Send a Cow's Director of Research and Impact, has noted several benefits from developing the theory of change and the measurement framework. These include helping staff and trustees buy into what is measured and how, identifying areas of commonality in how their programme is implemented in different countries, and then articulating what Send a Cow does and the impact it has. While it is too early to attribute any changes to the improved framework, individual donors and funders have commented on the clarity of the impact report and its robustness. One donor observed that very few NGOs are able to demonstrate their impact in a way Send a Cow can.

NEXT STEPS AND IMPLEMENTATION

Analysing data

Once you have worked through the four pillars of your measurement framework, you can start to put it to use. The point of the whole exercise is to help you understand how well you are doing, and how you could improve your work to help more people, more effectively. Analysing the data you collect allows you to test whether your theory of change works in practice—whether you are approaching your work in the best way, or could do better.

Used well, your measurement framework enables you to prove to staff, funders, and beneficiaries that 'by doing X, we achieve Y'. But to get to this point, you need to spend time analysing the data you have collected to draw meaningful lessons and conclusions.

How you analyse the data depends on your context and the type of work you do. Typically, you should aim to answer three questions.

1. Does your quantitative and qualitative data support your theory of change?

Does it suggest that the causal mechanisms that underlie your theory of change work in the way you expect them to? This leads to further questions:

- Are there any causal links in the chain that do not work as you expect them to?
- Does your service work the same way in different places or with different people? Can you identify reasons for any differences in the intermediate and final outcomes?
- What factors facilitate your theory of change and what factors impede it?
- Does the data point to any untested assumptions about how what you do leads to change?
- How much difference do you make (ie, what is the scale of your impact)?

Qualitative data, such as interviews with staff, beneficiaries and other stakeholders, is good for explaining why change happens—the causal mechanisms at play. It tells you more about *how* you have an impact, which quantitative data generally does not do. But qualitative data can be more susceptible to bias, and it is often difficult to tell whether the experience of an individual is representative. Combining qualitative and quantitative data provides a richer picture of your impact than either can alone, and will help you understand the difference you make.

Answering these questions will help you identify how you might be able to increase your impact. You may want to change some activities, stop those that appear to make little difference, or engage in new ones. As you think about shifting resources in this way, you not only need to think about becoming more effective, but also more cost-effective, which is different. Starting new activities might make you more effective, but the increase may not justify the cost.

2. How do your results compare with similar services or organisations?

Comparing your results with others ('benchmarking') can help you understand how you are doing in relation to your peers. This can offer insights into the level of success you can expect from your work, and suggest ways you might improve. But you need to be sure you are comparing apples to apples: are your results measuring the same

thing, in the same way, in the same context? For example, the difficulties in standardising cost-benefit and social return on investment analyses make it inappropriate to compare different services by these measures.

It is important to be aware of potential differences between your impact and that of a similar service that are out of your control, such as a different context or the effects of chance. But it is useful to know if your results differ from the norm, assuming there is such a norm. If you are within the norm you may want to think about how to improve. If you are outside the norm you will want to take action—either your results will be significantly worse, in which case you should stop or change what your do, or your results will be better, in which case you may want to start gathering a higher level of evidence.

3. Are your results improving over time?

This is the most important question to answer. You want to get better. It is useful to be able to show funders and other stakeholders that you can learn and improve. The challenge is taking account of factors that influence your impact but are outside your control.

For example, it might be difficult to account for changes in macroeconomic conditions, such as employment, that affect your efforts to help people get jobs. Again, this is where you can refer to your theory of change. Can you show you make a difference with the metrics that are in your control? Do these improve over time? Can you explain why those improvements do not lead to goals further downstream? Can you do anything about the factors that influence those downstream goals, such as collaborating with others?

Box 2: Analysis checklist

Make sure you have the right analytic capabilities. You may have these capabilities within your staff, but you could supplement them with external advice if necessary. External support will likely also provide credibility and verification, depending on the quality of advice and the reputation of the advisor. Beware of consultants who simply tell clients what they want to hear. This may provide a short-term boost but will not help you build a high-performing and high-impact organisation.

- Focus on your most important questions first. Start with what you think are the most important indicators of your success and analyse the data you have to help explain those indicators. It can be useful to frame your analysis as a set of research questions.
- Use all the information available. This might include quantitative and qualitative data, routine monitoring data, and in-depth evaluation data. Creating a clear picture of your impact is like putting together a jigsaw puzzle—you want to include all the pieces of data you have.

Be objective and clear-headed. Whoever conducts the analysis should take a cold look at your impact and not be biased. The purpose of analysing your impact data is to understand your impact and your ability to control it. Be aware that chance and causes outside of your control will affect your impact, and that you may not be as effective as you think.

Barriers to successful implementation

A number of barriers may hinder your progress in implementing NPC's four pillar approach. We outline them below so that you can think about how to minimise them before developing your framework.

Organisational culture or leadership not committed to improving

There can be a tension between measuring to show you make a difference, and measuring to learn. Supporters of a particular intervention or organisation often want measurement to show it in the best light, but this could undermine the goal of learning. We believe that learning from measurement should trump a desire to look good and should be reflected in a mission-driven culture and strong leadership.

Some charities, even those founded on sound principles and staffed by well-intentioned and competent people, are not committed to maximising their impact. They may be stuck in a routine and find themselves struggling to survive rather than think about ways to improve. Adopting NPC's approach may help to shift culture, but it cannot create an impact-driven charity in isolation. It needs to be supported by the other steps of the *Journey to greater impact* mentioned in the introduction.

Whatever the size of the charity, the success of a measurement framework depends on a logistical lead as well as senior champions. In our experience, the organisations that struggle most with our approach are those that do not single out one person to be responsible for the measurement framework and its results.

Lack of staff skills

The skills needed to measure your impact typically differ from those needed to create your impact. Evaluators need an objective and detached perspective. Bringing about change requires confidence and strong commitment to a cause—the opposite of detachment. While large organisations can have a team of analysts and researchers to assess impact, small charities can find it hard to combine both perspectives in their staff.

An organisation also needs the skills to make use of data that is collected. Creating and implementing a measurement framework is only half the journey. The other half is analysing and using the information to change and improve services. This typically requires some analytic capabilities.

Instability

If the organisation or environment is rapidly changing, creating an elegant and detailed theory of change and a measurement framework may be a waste of time and energy. Measures collected in one place and time cannot sensibly be compared to measures collected in a different place or time. However, even in the midst of change, a high-level theory of change may provide a sense of purpose and stability and help an organisation remain true to its goals. In such an environment, creating short-term feedback loops is critical to developing a programme or service that really makes a difference.

Complexity of advice on evaluation

The plethora of advice on how organisations should measure their impact can be confusing. Some sources advocate only the most robust methods, such as randomised control trials; some provide tools to help track subjective views on what is working well; others recommend aggregating impacts into a monetary value. NPC's approach provides guidance on how to make sense of all of the evaluation options available, but knowing which tools to use when may still be unclear. You may need to consider external help to implement the approach.

CONCLUSION

This report provides guidance on how charities can track their impact so they can improve, but also demonstrate their impact to others, including funders. We have touched on a number of theoretical and technical challenges and provided guidance on how to address them. Mostly, we have provided principles to follow and ideas to try, which our clients have found helpful (see the 'clients and partners' section of our website). It is up to you to try this approach yourself, and identify what does and does not work.

What this guidance does not do is make the case for charities to start this process in the first place—we assume this already exists in your organisation. The vast majority of the people we work with want to make a difference. It is up to chief executives, senior managers, and trustees not to waste this goodwill and energy by having staff continue to do things that have limited impact, no impact, or even a negative impact. The only way to do this is to find out what is working well and what is not. As one of our clients put it, *'Why would you want to work in the dark?'*

We leave you with some final tips for implementing NPC's four pillar approach. Bear these in mind throughout your measurement journey, to ensure you get the most from your efforts and use your data to get better.

Box 3: Making the most of NPC's four pillar approach

- **Decide what is most important to your organisation and try to measure that**. Do not be too easily pushed into what funders want measured or what is easy to measure.
- **Look at possible negative impacts of your work**. This is the right thing to do, and you want to be prepared should someone challenge you.
- **Think about where you want to be in the future and what data you need to help you get there**. It takes time to build a high-impact organisation. Data on your impact will help show you the way.
- **Enemy number one is poor quality data**. Collecting it serves no purpose but consumes resources.
- **Keep your measurement approach and data collection as simple as possible**, but collect sufficient data of good quality to illuminate whether your theory of change is working as you anticipate.
- **Know the limits of what you can and cannot do**. Consider investing in training your staff and seek external advice where necessary.
- **Use data you collect to inform your practices**. It is the only way to get better.

Tell us what you think

NPC's mission is to help charities and funders improve the lives of their beneficiaries. We seek to help them become more effective and efficient, through a range of services, advice and research. As we are on our own journey to improve, we would very much like to hear your views about our approach, and what more we could do. Please contact info@thinkNPC.org with your comments.

APPENDICES

Appendix 1: Client experience of using NPC's approach

The crucial test for any theory, idea, or hypothesis, is whether it works in practice. NPC's approach is not the only way to develop a measurement framework. Many charities do so without first going through this approach.

But we have some evidence as to how well it works. As well as using this method ourselves, we provide training in implementing this approach and we provide consulting services to charities and funders who ask for further assistance. As of April 2014, NPC has completed consulting projects for 16 organisations that have contracted us to help them develop a framework using this approach in full. We are currently working with several more organisations and have worked with several other organisations using only parts of this approach.

Of the 16 organisations that commissioned us to develop and implement measurement frameworks using this approach:

- Nine, including Mind, Business in the Community and Send a Cow, are using some or all of the framework to assess and/or report their impact.
- Five, including Shaw Trust and Breakthrough Breast Cancer, found the approach of working through a theory of change very useful and have used it to clarify their thinking about their impact. They found the measurement recommendations helpful but are not yet routinely producing impact data around the framework.
- One tested the framework but had difficulties embedding it into its work. We will be revisiting the framework with the charity to address the challenges.
- One found developing the theory of change very useful but did not find the development of a measurement framework useful.

The experience of four organisations that have used this approach—Mind, Business in the Community, SolarAid, and Careers Development Group—are featured on our website at www.thinkNPC.org.

We have also taken approximately 100 individuals from different charities through the four pillars approach at day-long public workshops. In their feedback forms, 83% (69 out of 83) said they 'will use what they have learned', and 59% (44 out of 76) said they 'will makes changes in their charity as a result'. We have also supported a similar number of people in workshops held for particular organisations.

Appendix 2: Levels of rigour

Social scientists generally agree that certain types of evidence are better than others. Accordingly, some have created scales or levels of evidence that represent a range from the least to the most robust measurement approaches.[14] Two such scales used in the UK charity sector are:

- **SRU's Standards of Evidence**—a five-point scale used by Project Oracle, the Big Lottery Fund's Realising Ambition programme, Graham Allen's reports on early intervention, and adapted by NESTA for use in social investment.[15]
- **The Maryland Scientific Methods Scale**[16] —another five-point scale originally used to evaluate the evidence in crime prevention programmes, which has since been applied more widely. It has been used by US and UK government departments.

Both these scales provide a guide as to what makes weak and strong evidence. Each has five levels of evidence, though the levels are not the same for both scales. They start with a weak level of evidence, namely a theory of change for SRU's Standards of Evidence and correlation for the Maryland Scientific Methods Scale. They both categorise Randomised Control Trials as the most robust evidence. Both scales apply to the rigour of quantitative methods of assessing impact, but others have tried to apply standards to qualitative research as well.

Of course, these scales are simplifications, and do not provide a definitive assessment of the quality of evidence. There are areas of disagreement as to when certain types of evidence are most appropriate. Sometimes a good quality study using one level of evidence may be better than a poor quality study using a higher level of evidence. For example, to an expert, a robust before-and-after study with no control group may provide better evidence than a poorly designed and executed randomised control trial. Context and information specific to the particular evaluation in question are important.

Appendix 3: Common methods of showing cause and effect

This appendix describes common methods available to charities using either qualitative or qualitative data or both to demonstrate causality.

We have grouped the methods according to which of the four approaches they best fit. However, as noted in Step 3, the approaches are not mutually exclusive and all of the methods could be used for any approach.

Statistical approach

- *Correlation*: This tests whether participation in a service or programme is associated with the intended outcome, such as whether young people who attend a training course are likely to get jobs. But correlation does not mean causation: the course may simply attract motivated young people. Without a control or comparison group you cannot rule out alternative explanations.
- *Before and after comparisons (or pre-test post-test)***:** These may be surveys filled in by a participant, or some form of objective measures, such as measures of health, employment, housing status, before (or at the beginning of) a programme and then again at the end of (or shortly after). This method is used to see if any change has occurred between the two periods. However, this change does not necessarily demonstrate causality because there may be other factors involved.
- *Observation***:** Staff or researchers observe changes in people, policies, etc. and look for patterns in those changes that suggest a cause and effect relationship. Data used to assess change can come from almost any source, such as staff observations or government administrative data. Observation studies are distinguished from experimental studies as, in the latter, the charity deliberately manipulates who participates in the charity's intervention or service and who does not. In the former, the charity simply observes the changes in participants.

Experimental approach

The credibility of this approach depends on how well the control or comparison group is selected. Benchmarking against the outcome measures of similar organisations is a way of establishing a counterfactual, but there may be important differences between who you work with and who other organisations work with that will explain any differences. Therefore you cannot pick any comparison group, such as a national average, and always expect that to be credible. The most credible methods of creating a counterfactual are:

- *Randomised Control Trials (RCTs):* While an RCT provides strong evidence of the impact of a service at one place and one time, it does not necessarily follow that the same service will have the same or similar impact at another place and/or time. Thus the strongest evidence comes from several RCTs being undertaken in different places and times and producing similar results. See Appendix 4 for alternatives to creating control groups by randomising who receives and does not receive a service.
- *Quasi-experiments:* It is sometimes possible to compare the outcomes of service participants with non-participants without randomly assigning people to one group or the other. In quasi-experiments typically the comparison group is created from data sets—such as government data—that contain information about non-participants as well as participants. Statistical methods such as regression analysis can be used to estimate the likely effect of participation on the outcomes being measured. This category also includes quasi-

experiments where the control group is not selected by random assignment but is selected to closely match the intervention group. For example, the Justice Data Lab compares the reoffending rate of a group of a charity's service users with the reoffending rate of a control group made up of similar ex-offenders that the Ministry of Justice identifies from its records. The ex-offenders in the control group are picked based on how closely they match the charity's ex-offenders in terms of known risk factors such as age, gender, their history of offending, and so on.

- *Natural experiments:* A natural experiment uses differences in where, how, and with whom a programme or service is implemented to identify causes. For example, suppose a charity has a programme in a number of schools in London, but the programme is implemented in three different ways depending on the school. Comparing data from the different schools, as well as similar schools that do not implement the programme, can allow inferences about whether the programme makes a difference and how, where, and when it is implemented makes a difference.

While RCTs and quasi-experiments tend to be costly (because they involve collecting data from people who do *not* participate in your service as well as those who do) the Justice Data Lab is currently free for charities to use and so provides credible evidence at a low cost.

Comparative approaches

- *Case studies:* case studies are a record of research into the development of a particular person, group, or situation over a period of time. Comparing multiple case studies with each other can help identify patterns of conditions and causes that together appear to make a difference in outcomes.

Theory-based

- *Individual case studies:* Individual case studies, done well, can provide strong evidence and are very useful in illustrating and explaining the causal mechanisms in your theory of change and showing impact when it is not possible to create a control group. But they are also susceptible to bias, such as when the 'best' cases are implied as being typical. For example, a charity that seeks to influence policy may analyse the results of a successful campaign and show how it managed to persuade policymakers. But if it ignores cases of unsuccessful campaigns, it is not only misleading in its impact—and may be wrong about its influence—but it may not learn how to improve.
- *Self-reported change:* Beneficiaries or service users read a question and select a response by themselves without researcher interference, usually giving a view on whether the service has made a difference to them or not. While often compelling, these responses can be subjective and unreliable. It is often difficult for beneficiaries (or other stakeholders) to judge whether any change is attributable to a particular intervention or programme, and respondents may want to give pleasing messages. Data from other people—such as a practitioner or a family member—can be used as an additional point of view on whether change has been achieved and why.

Appendix 4: Control groups

The Cabinet Office paper Test, Learn and Adapt*,* by the Behavioural Insights Team, Ben Goldacre, and David Torgerson explains why randomised trials of policy using control groups are powerful, and how to do them.[17] Using examples of policies that should have worked in principle, the paper explains how they turned out to be harmful when finally tested.

Randomised Control Trials are seen by some as the most robust way to measure impact, but they are not suitable for all contexts and situations. A control group may be less important if behaviour has been static for your user group for several years and you can reasonably infer that any change is due to your programme. This approach may be relevant to some more extreme or difficult user groups, such as the long-term unemployed or people with serious disabilities, but even in these instances having a control group provides stronger evidence.

Creating a control group throws up several challenges: some situations are harder to randomise than others, for instance, when targeting the hardest-to-reach or specific geographic areas. Under these circumstances, you may have to show your attribution in other ways, such as by the clarity and logic of your theory of change and collecting rigorous qualitative data.

Even when it is possible to create a control, it is typically difficult and costly to identify, recruit, and track people who do not participate in your programme. It is often hard enough to do this with people who are in regular touch with your organisation, let alone people who are not. Many charities find it difficult to justify the cost.

There are more straightforward ways of setting up a control, such as:

- **Using people on the waiting list**. If your service can only take a restricted number of users and the demand exceeds that number, you can end up with a waiting list of possible future users. These people have an interest in staying engaged and are likely to be open to being surveyed while waiting for their placement. To be comparable, however, the reason for being on the waiting list needs to be random, rather than other reasons, such as having a lower priority of needs.
- **Giving similar users different levels ('dosage') of a service**. You can do this if you run several programmes with different degrees of intensity but for similar people. For example, a charity that works with young people who are at risk of becoming homeless can compare the outcomes of its intensive counselling service with its light touch drop-in-service to see the difference. It is important to ensure that the two groups share similar characteristics (eg, have similar needs) so you can compare apples to apples.
- **Comparing your results to national/local authority data for similar people/services**. If you work with many users whose characteristics are similar to the general population, you can use statistics published on the latter as a comparison group. This can be relevant for large prevention programmes where users come from the general population and are not selected because they are different. Useful data sources for national and regional statistics include UK National Statistics, government departments and the Understanding Society Survey.[18] At a local level, you may be able to get comparable data from local authority websites. Some government departments also release local statistics (eg, Metropolitan Police Crime Mapping[19]) at borough and ward level. Note that making comparisons to a general population (national or local) will probably be meaningless if the people you work with are not representative of a general population.

The examples above will not be appropriate for every organisation and will depend on the context in which you operate. Even so, at best, a control group comparison is an estimate. It is an alternative scenario which is the least likely to be misleading or biased. There is no such thing as a perfect counterfactual in the real world. What is important is to balance the highest level of rigour with what is feasible in practice. It is important to be open about the limitations of the analysis when it comes to reporting results.

Appendix 5: Sampling

Sampling can be an efficient way of getting data about a large group (population) from a small group (sample). It allows you to make claims about your impact based on a sample of beneficiaries, not all of them. This helps keep costs down.

In choosing a sample you generally need to consider these four questions:

- What do you want to know about the population?
- How representative is your sample of your user population?
- How large should the sample be?
- What is your expected response rate?

What do you want to know?

Quantitative research: You need to use a quantitative sample if you want to make generalisations about a large population. In this case, it is best to sample respondents randomly if possible (see below). Achieving a good response rate is important as people with certain characteristics or stronger views may be more likely to respond to a survey. Sample size is also important. If your sample adequately represents the group it is drawn from, a general rule of thumb for a good minimum number of cases is 50. If you have less than 50 cases, or you know your sample is heavily biased towards a particular type of respondent, it is best to acknowledge this in your reporting, and avoid using percentages in reporting the findings as they will have lost meaning.

Qualitative research: The aim of qualitative sampling is often to cover the full breadth of experience or understand the process by which impact is achieved rather than produce statistically representative estimates of the population. The response rate is less important as long as the full range of key characteristics is represented in the sample. But if you are sampling to get the full range of experiences, you should not then make claims about how those experiences are reflected in the population.

How representative is your sample?

The main ways of selecting a sample are: random, quota, purposive, and convenience. Each has their pros and cons: the right choice depends on what you are trying to find out.

Random and quota (for quantitative research): When you want to make claims about numbers, you want to select the sample randomly. It is the random selection that allows you to make estimates about the population from which the sample is drawn. If you do not select randomly, you may create biased results and those estimates become invalid. However, in order to make sure your sample is representative of your user population, you may want to select a random sample within subgroups. For example, if your user population consists of 70% girls and 30% boys, or 10% in each of ten locations, then for quantitative sampling your sample should probably have similar proportions. Depending on the sampling technique you choose, ensure that each sub-group is represented with the right proportion. If the sample population does not map well to the larger population you are interested in, you can weight data for sub-groups in the sample to make it more reflective of the larger population. Quota sampling is where you stop contacting or approaching people to take part in your survey once you have reached target numbers per category (eg, number of girls), whilst random sampling is where everyone in your sample is contacted the same number of times—the former has more response bias as those less likely to respond (eg, with less strong views on a topic) are less likely to have been included in the survey.

Purposive (for qualitative research): When you are collecting qualitative data, it is more important to collect the *range* of views and experiences rather than the exact proportions (for example 10 fathers and 10 mothers using a children's centre, even though the children's centre is used predominantly by mothers). You don't have to be as careful about getting the same proportion as in reality.

Convenience (for qualitative research but sometimes used in quantitative research): Convenience sampling is what it says. You pick the most convenient sample, such as the people you are in most contact with. Sometimes this may be necessary for cost reasons or where there are no other options but generally it is not very rigorous and avoid it if you can. Any analysis, especially quantitative, should acknowledge the sampling approach and the potential bias associated with it.

How large should your sample be?

A number of factors influence the most efficient (ie, smallest) sample size. For quantitative research, as a rule of thumb, you want to end up with a data set of 50 at a minimum.[*] This will require sampling more than 50 expecting that some in the group you sample from will not respond and others will drop out etc. If your user population is less than that, you can still do quantitative analysis but your results are then more likely to be due to chance, and it will be harder to make a persuasive case that your service makes a difference. In this case good qualitative data will be an important addition to your measurement framework.

For qualitative research, a common sample size is between 10 and 50 interviews. As a general rule of thumb, little new data is likely to be obtained through more than 50 interviews. The minimum sample size for your needs will depend on the variability of experiences you expect in your population.

What is your expected response rate?

For quantitative research you want to minimise the likelihood the data may be biased (eg, in a survey those with stronger views may be more likely to respond). Thus you want to get data from as many in your sample as possible. If necessary, you may want to make a concerted effort to reach those in the sample who you missed in the first round of data collection to make sure they are not, as a group, different in significant ways to those who you do have data on.

Appendix 6: Measurement methods and tools

There is a large number of tools that can be used to collect data to inform the impact of your charity. Here we divide these tools into quantitative and qualitative tools. There is no single source for finding appropriate measurement tools. At NPC we use a number of different sources to identify appropriate tools (see Appendix 8), and sometimes we help develop new ones. The Inspiring Impact programme (www.inspiringimpact.org) is launching a website in June 2014 that will provide a link to common tools used by charities.

Here are the most common types of measurement tools.

Quantitative tools

Surveys and questionnaires

Surveys and questionnaires are a common form of measurement tools because they can be targeted to the right population and customised around the outcomes you want to measure. They assume that people will be honest in their answers and sufficiently aware of the outcome to provide credible answers. These assumptions do not always hold. For example, if you ask people at the beginning of a course to assess their own competency in a particular skill (such as parenting) and then train them in that skill, you might find a counter-intuitive pattern. People are likely to over-estimate their own competency at the beginning of the course, then score themselves lower once they find out what they didn't know. On its face the self-assessment data would show the training led to a decrease in competency. There is no simple solution to this problem: triangulation may be part of the answer, but it is important to collect data so you are aware of this problem in the first place.

We recommend using existing surveys that have been created by researchers, but if you need to develop your own question (or questionnaire), it is best practice to develop questions / questionnaires from existing material—selecting questions that are most appropriate to what you need to ask. There is a large resource of government surveys that have a wide variety of questions that you can access. You can search for surveys and questions at

[*] You can use an online sample size calculator to help you use your preference for confidence intervals and levels, and determine the right sample size: http://www.surveysystem.com/sscalc.htm. Some basic statistics knowledge will be necessary to be able to use this tool.

the UK Data Service's website at http://ukdataservice.ac.uk/get-data.aspx. If you need to develop your own questions, try to:

- **Be brief:** avoid going over 20 words, remove unnecessary words.
- **Be simple:** avoid complicated words, confusing language.
- **Be specific:** avoid words such as 'often' or 'usually'.
- **Be objective:** avoid leading questions. Include a range of answers and consider a 'don't know' option.
- **Use your common sense:** ask yourself, 'Will the respondent understand the question?'
- **Test questions with respondents where possible:** Ask the respondents what they think the question is asking. This will help you assess whether it is getting at the right outcome.

You may also want to take advice from a researcher.

Psychometric scales

Psychometric scales are a particular type of short questionnaires. They are designed by psychologists and sociologists to measure subjective feelings, beliefs and attitudes, such as self-esteem or empathy. They usually include a series of statements, with respondents indicating on a scale the extent to which they agree with each statement. Psychometric scales are the most robust way of demonstrating change in soft outcomes, such as self-esteem or attitudes. They are designed to be objective and unbiased, and are rigorously tested for validity and reliability. Importantly, they cannot be altered or changed as every statement in the scale is given a value, which is added together to produce a total score. To decide on a scale, you should read all questions or statements in the scale to understand whether:

- the questions/statements reflect the outcomes that your programme aims to achieve;
- it is the right length (some scales are long);
- it has been tested with the right age group; and
- the language is appropriate for the culture you are testing in.

If the scale does not satisfy any of the points above you should look for a more appropriate scale or design your own questions.

NPC has a on-line tool that measures the well-being of young people via the combination of a set of psychometric scales.

Case work tools

Case work tools are used on a one-to-one basis where a case worker or other member of staff assesses progress with an individual on a number of areas in that person's life. Well-known case work tools include the Outcomes Star family of tools.[*] These can be used at an individual level to identify the needs of the client group and to measure progress or distance travelled for individuals. They are popular with practitioners because they help them deliver services as well as measure progress. They can also be aggregated to show change at a service level. The tools are, by nature, subjective, and are designed to be used on a one-on-one basis with a client. Because they are filled out together or self-rated, there may be a bias on the part of both the client and the practitioner to want to show progress over time: this can result in false positives where the client or case worker judges they have moved up the scale when not much may have changed. In addition, the steps of progression are typically changes in the client's attitudes, feelings and behaviour that practitioners and clients believe represent progress, rather than 'objective' indicators backed up by research. For this reason, they are not

[*] http://www.outcomesstar.org.uk/

considered to be as robust as using clinical tools or psychometric scales to measure soft outcomes, although some of the Outcome Star family have gone through testing.[*] Data produced from case work tools may show changes in outcomes, subject to the caveat above, but without a control group do not show the impact of a service. Hence, organisations that use such tools to show the difference they make need to use supporting data to back up claims made.

Statutory or administrative data

As described above, statutory agencies can be a source of data and tools. Government data on individuals, communities, schools, businesses, and so on, has already been collected by statutory agencies. This presents two problems for a charity: assessing the quality of the collection tool and data, and accessing the data. Do not assume that government data is high quality: some government data is better quality than others. Furthermore, government data is typically difficult to access.

NPC is working to improve charity access to government data. One such project is the Justice Data Lab. Charities that are working to reduce reoffending can request aggregate reconviction rates of a cohort of people that the charity has worked with from the Ministry of Justice (MoJ). MoJ will also provide the aggregate reoffending rate for a matched sample control group. This is currently available at no cost to the charity, making it a very efficient and cost-effective means of assessing impact.

Qualitative tools

Interviews

Qualitative interviews are frequently used in social research. They differ from questionnaires in that the answers are not pre-categorised as they are in questionnaires and the interviewee is given the space to expand on their answers. Interviews are normally used to give more detailed information on a particular topic and to explore the topic from a person's own viewpoint. Qualitative interviews are more appropriate to understand how and why change might have happened and are generally not easy to use as measures of impact.

When interviewing you should:

- prepare a topic guide rather than a questionnaire;
- use open questions (such as 'How ?', 'Why ?', 'What way ?');
- avoid suggesting answers or biasing responses;
- approach sensitive topics carefully, taking into account impact on interviewers; and
- ensure all interviewees are asked the same question while allowing the interviewer to probe the interesting answers.

Observation/ethnography

Observation involves the systematic observation and recording of behaviours and interactions of people in their environment. Observation is used to generate detailed descriptions of certain behaviours or traits and can be useful in recording information that is otherwise difficult to obtain with other research methods. When carrying out observation, it is important to:

- have a theory about what you are observing;
- decide on your focus;
- systematically record what you have observed; and

[*] http://www.outcomesstar.org.uk/research-and-briefings/

- use analysis to identify themes and patterns of behaviour to generate any conclusions.

Like interviews, observation is best used to understand a topic or issue in detail and may not be the most efficient tool to measure impact quantitatively, though it can be done.

Focus groups

A focus group (or discussion group) is a type of research method where a group of people are asked about their thoughts, attitudes and beliefs towards a topic of interest—whether that be a product, service or social issue. Questions are posed to the entire group and people share their thoughts in an interactive manner. Notes are taken at focus groups and later analysed to look for trends and themes. Focus groups are very useful to understand why a service leads to change for people, but because of the subjective nature, the possibility of bias (focus groups can be dominated by the strongest voices), and because of the small numbers involved, they are not appropriate as quantitative measures of impact.

Ad hoc data and feedback

Ad hoc data and feedback arise from day-to-day interactions with your beneficiaries or other stakeholders. These interactions may contain stories and feedback on how your work may have positively impacted someone or something. Some of these anecdotes may be hard outcomes, such as a change in legislation or someone moving into employment, and these can be recorded systematically: other anecdotes may be 'soft', for example, someone telling you how their confidence has improved or how they made new friends on your course.

Such feedback is useful and is one piece of evidence that can be used to test your theory of change, but it should not be used as the only or main source of impact data.

A spreadsheet can be used to capture systematically any ad hoc data on your impact. Recording e-mails or oral comments, texts, and so on, can help an organisation see trends in these stories and build up a picture of impact over time.

Example of an ad hoc data spreadsheet

Project	Ad hoc data	Outcome/Goal	Impact	Attribution	Source	Date
Employment skills course	X told us that the module on CV writing helped him get an interview	Support young people into employment	High	Some	During the session	June 2013
Mentoring scheme	X told us that 'the scheme helped her relationship with her parents'	Support young people with social and emotional capabilities	High	Some	In an e-mail	August 2013

Appendix 7: Other approaches to measuring impact

This report is designed to help charities adopt an 'impact approach' to what they do. The four pillars approach is a step-by-step guide for doing this. We do not advocate a particular evaluation methodology, a particular framework or a particular tool. Charities use many approaches to try to assess social impact.[*] The four common approaches below are not direct substitutes for each other, and each has its own uses. The four pillars approach fits into the first category.

- **Linking goals and objectives to activities**: NPC's approach to measuring impact revolves around a theory of change to describe the causal mechanisms by which an organisation expects to achieve its goals. Other common models that use a causal chain approach are logic models and log frames (based on the Logical Framework Approach[†]). Logic models describe how an organisation's inputs and processes relate to its outputs and outcomes. Log frames are widely used in the development sector as a project or programme management tool, and link activities to outputs, purpose, and goals, along with relevant assumptions. Logic models and log frames are similar to theories of change as they also focus on the causal relationships between activities and outcomes. We feel that logic models and log frames are too linear to properly reflect the way charities create change or to illustrate the combined impact of a charity's different activities. Whichever model is used, charities should ensure that the causal mechanisms that make up the model are not only plausible in theory but actually happen in practice.
- **Assessing organisational effectiveness**: Some tools and models that are described as measuring social impact instead aim to assess organisational effectiveness. Hence these assess the *likelihood* that an organisation or programme does or will have a positive social impact, rather than the *degree* or *amount* of impact. Many approaches in The Foundation Center's TRASI (Tools and Resources for Assessing Social Impact) website fall into this category, including NPC's own charity analysis tool *The little blue book*, or Charities Evaluation Service's PQASSO tool.
- **Measuring 'distance travelled' of beneficiaries**: Two groups of these kind of tools are often used by practitioners working one-to-one with individuals to help them improve their lives: clinical tools that describe someone's mental or physical health; and tools that combine a road map for improving and a means of tracking improvement or distance-travelled. The best example of the latter is the Outcomes Star family of tools. Because this category of tools is practitioner-led, they tend to be easy to implement and useful to staff. The risk is that as they involve subjective judgements on progress—depending on how they have been developed and how they are implemented—they are susceptible to producing biased data.
- **Assessing economic impact**: Success is not only about the effectiveness of a service, but also its cost-effectiveness. A high-impact, high-cost service may be more cost-effective than a low-cost, low-impact service—the latter could provide "more bang for your buck". There is always interest in identifying, measuring, and reporting on cost savings and efficiency gains, and converting social impact into monetary terms. Approaches that do this, such as cost-benefit analysis and Social Return on Investment, can be very useful, but done poorly the results can be misleading. Estimating the financial or economic value of a charity's impact typically requires first collecting robust impact data, then converting some or all of it into monetary values. NPC's approach can be used for the first step.

[*] A good article that covers the difficulties in measuring social impact can be found at: http://www.ssireview.org/articles/entry/measuring_social_value.

[†] The Logical Framework Approach is a management tool mainly used for the design, monitoring and evaluation of international development projects.

Appendix 8: Other resources

A number of websites and publications can help you think about how to identify data sources and develop tools.

General

- The resource page of the Inspiring Impact programme includes a list of measurement tools and systems: http://inspiringimpact.org/listings/
- TRASI has a database of different approaches to impact assessment: http://trasi.foundationcenter.org/
- Ritchie et al. (2003) *Qualitative Research Practice—A Guide for Social Science Students and Researchers*. Sage Publications Ltd.
- NPC's *A journey to greater impact*: profiles six charities that radically improved their approach to impact measurement: http://www.thinknpc.org/publications/a-journey-to-greater-impact/
- Improving your evidence: a joint NPC and Clinks programme to help charities working in the criminal justice sector: http://www.clinks.org/support/evaluation-and-effectiveness
- Better Evaluation website: http://betterevaluation.org/

Theory of change

- A website dedicated to theory of change: http://www.theoryofchange.org/
- NPC's report on *Theory of change*: http://www.thinknpc.org/publications/theory-of-change/
- The Innovation Network's logic model workbook: http://www.innonet.org/client_docs/File/logic_model_workbook.pdf

Outcomes

- NPC's report on *Mapping outcomes for social investment* designed to help social investors, and those seeking social investment, to embed a robust approach to impact in their work: http://www.thinknpc.org/publications/mapping-outcomes-for-social-investment/

Measurement tools

- A good website to look for measures and indicators: http://wilderdom.com/tools/ToolsSummaries.html
- A website where you can find outcomes and standardised scales relevant to a range of human welfare services: http://www.performwell.org/
- NPC's *Journey to employment (JET) framework* designed to help organisations working in youth employability to understand and measure their impact: http://www.thinknpc.org/publications/the-journey-to-employment/
- A list of the most common psychological scales and questionnaires: http://www.ull.ac.uk/subjects/psychology/psycscales.shtml
- Innovation Network's Point K website with a list of useful resources: http://www.innonet.org/

Surveys and questionnaires

- Bradburn et al. (2004) *Asking questions: the definitive guide to questionnaire design*. Jossey-Bass: San-Francisco.
- Survey question bank: a website where you can research widely used surveys and single questions by theme or using key word searches: http://surveynet.ac.uk/sqb/

Reporting results

- How to communicate your results: Hedley, S. et al (2010) *Talking about results*. New Philanthropy Capital. http://www.thinknpc.org/publications/talking-about-results/

REFERENCES

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

[2] Van Vliet, A., Kail, A. and Baumgartner, L. (2013) *Funding impact: Impact measurement practices among funders in the UK*. New Philanthropy Capital.

[3] Harlock, J. (2013) *Impact measurement practice in the UK third sector: a review of emerging evidence*. Third Sector Research Centre Working Paper 106.

[4] Rickey, B., Ní Ógáin, E. and Lumley, T. (2011) *A journey to greater impact: six charities that learned to measure better*. New Philanthropy Capital.

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

[6] Copps, J., Kail, A., Plimmer, D., Ní Ógáin, E. and Harries, E. (2014) *The Journey to EmploymenT (JET) framework*. New Philanthropy Capital.

[7] http://www.slideshare.net/pkebel/creating-your-theory-of-change

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

[9] Befani, B., Davies, R., Forss, K., Mayne, J., Stame, N. and Stern, E. (2012). *Broadening the Range of Designs and Methods for Impact Evaluation*. Department for International Development Working Paper 38.

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

[11] Funnell, S. C. and Rogers, P. (2011) *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models*. John Wiley & Sons, Inc.

[12] http://www.clinks.org/sites/default/files/IntroductionToSampling_0.pdf.

[13] De Las Casas, L., Ní Ógáin, E. and Svistak, M. (2013) *Blueprint for shared measurement*. New Philanthropy Capital.

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

[15] http://dartington.org.uk/projects/what-works-evidence-standards/

[16] https://www.ncjrs.gov/pdffiles/171676.PDF

[17] https://www.gov.uk/government/publications/test-learn-adapt-developing-public-policy-with-randomised-controlled-trials

[18] https://www.understandingsociety.ac.uk/

[19] http://maps.met.police.uk/?areacode=00AG