



What works?  
Eight principles for  
meaningful evaluation  
of anti-prejudice work

Colin Duff and Carol Young

Coalition for Racial Equality and Rights



Equality and  
Human Rights  
Commission

# About this publication

## What is the aim of this publication?

This guidance has been developed to help organisations measure the effectiveness of interventions that have been designed to tackle, prevent or reduce prejudice, discrimination, identity-based violence or harassment. Presented as a set of eight principles, the guidance should be used as a planning tool to identify whether a new or existing intervention is working and where there may be opportunities for improvements. Application of these principles should allow users to achieve a minimum standard of evaluation of anti-prejudice interventions.

## Who is it for?

The guidance is intended for project coordinators, policymakers, evaluation and monitoring managers and funders who develop, implement or evaluate anti-prejudice interventions. This can be across different sectors, including third sector organisations, statutory bodies, central and local government and funders.

## What is inside?

**Part 1** provides a background to the guidance, setting out the steps that were taken to produce and identify a minimum standard of evaluation for anti-prejudice interventions

**Part 2** sets out an overview of the benefits of evaluation to an organisation

**Part 3** highlights the main barriers and challenges that arise in an organisation's ability to conduct a minimum standard of evaluation

**Part 4** introduces eight principles that have been designed to help organisations achieve a minimum standard of evaluation for anti-prejudice projects

**Part 5** lays out the first set of principles that organisations should apply before embarking on an intervention in order to make good decisions on what they wish to achieve

**Part 6** presents the next set of principles that ask what measures of success organisations will use to evaluate the success of an intervention

**Part 7** discusses the final set of principles that look at how organisations might use the data resulting from the evaluation to develop conclusions

## Why has the Commission produced this?

Our previous research has shown that there was a lack of anti-prejudice interventions that had been evaluated and shown to be working, and so there was little robust evidence that policymakers could draw on to prevent or respond to hate crime and other forms of unlawful behaviour that come about because of people's prejudice. We produced this guidance to support better evaluation and strengthen the available evidence on 'what works'.



# Contents

■ Executive summary .....	5
■ Introduction .....	7
■ Background .....	8
■ The value of evaluation .....	9
■ Barriers to evaluation .....	11
■ A minimum standard for evaluation .....	14
■ Designing the project .....	16
Principle 1.....	16
Principle 2.....	19
Principle 3.....	20
■ Planning the evaluation.....	22
Principle 4.....	22
Principle 5.....	25
■ Carrying out the evaluation .....	29
Principle 6.....	29
Principle 7.....	31
Principle 8.....	36
■ Conclusion .....	38
■ Useful resources .....	39
■ References.....	40
■ Appendix.....	42
■ Contacts.....	47

# Executive summary

**In July 2016, the Equality and Human Rights Commission published ‘Prejudice and unlawful behaviour: exploring levers for change’ by Abrams *et al.* This publication set out the current evidence on how prejudice relates to unlawful discrimination, harassment and identity-based violence.**

The report identified a need for better evaluation of which anti-prejudice interventions are effective and why, particularly for organisations that experience challenges and barriers to evaluating the difference they make. In response to this, the Commission asked the Coalition for Racial Equality and Rights (CRER) to develop guidance and training on how to evaluate interventions for organisations working to tackle prejudice, discrimination and identity-based violence and harassment.

The aim of this guidance is to build capability around how organisations can evaluate their interventions in a way that is proportionate and realistic. Better quality evaluation should yield better quality evidence on effective practice, which the Commission and others can rely upon to inform their work.

This guidance is presented as a set of eight principles. These principles support users to take an approach that brings together processes of intervention design, evaluation design and evaluation implementation.

Application of these principles should enable users to achieve a minimum or ‘good enough’ overall evaluation of anti-prejudice interventions, while still ensuring that the evaluation is sufficiently robust and transparent to allow confidence in the findings and their use in designing policy.

## Eight principles for evaluating anti-prejudice projects

### Designing the project

#### Principle 1

Our decision to make an intervention is based on a robust assessment and specification of the need to make an intervention.

#### Principle 2

We are clear about the difference we wish to make through our intervention.

#### Principle 3

We have reason to believe that the intervention we propose to deliver will produce that difference.

### Planning the evaluation

#### Principle 4

We are clear about the nature of the data required to demonstrate that we have made a difference.

#### Principle 5

We are clear about the methods we will employ to collect that data.

### Carrying out the evaluation

#### Principle 6

We know how we will analyse the data we collect to produce conclusions.

#### Principle 7

We know how we will use our conclusions.

#### Principle 8

We have assessed and committed the resources required to deliver the evaluation.

# Introduction

**Evaluation should be simple. It should be as simple as asking: ‘Is what we do making a difference?’ Unfortunately, evaluation is more complex than that. If it was that straightforward, then there would be no need for this guidance.**

Imagine you are an architect. It may not be wise to accept a brief that simply said ‘build a house’. You would want a little more detail and clarity than that. If you don’t have that clarity, then it’s quite likely that the house you build wouldn’t quite meet everyone’s expectations. Similarly, with evaluation, it may not be wise to accept a brief that simply instructed you to answer the question, ‘is what we do making a difference?’ Unless all the components are clearly and properly described, then the final evaluation may not meet anyone’s expectations, and may not even be an evaluation at all.

The aim of this guidance is to support you to incorporate ‘good enough’ evaluation processes into the planning stages of new interventions. An ‘intervention’ is any activity designed and carried out with the aim of tackling, preventing or reducing prejudice, discrimination, identity-based violence or harassment.

There are many forms of evaluation. This guidance is concerned with evaluation processes that allow you to measure the difference your work makes to tackling prejudice, discrimination, identity-based harassment or violence in Britain. Knowing if and how your work makes a difference is important. If your work isn’t making a measurable difference, why do it?

This guidance is intended to support improved self-evaluation; that is, evaluation that is planned and conducted by the same organisation that is delivering the intervention. It includes several lessons that can also be applied to commissioning an external evaluation. The guidance has been designed to be applied to the evaluation of project-level interventions. However, it should also be useful for partnerships managing or coordinating broader programme-level interventions.

The guidance is designed to support you to develop an overall approach that brings together the processes of intervention design, evaluation design and evaluation implementation. It is designed to be read as a whole document, and allow you to design an overall evaluation process that is ‘good enough’ to produce results you, and others, can be confident in.

Following the guidance from beginning to end may require you to have a ‘new’ project, which allows you the opportunity to build in each of the principles from the start. If you don’t have a ‘new’ project, or you don’t feel you have the time, skill or confidence to do it all, you can select one or more elements and use these to improve particular areas of your existing evaluation processes. It’s better to improve some of your evaluation than none at all.

# Background

**The Equality and Human Rights Commission wants to know ‘what works’ to tackle prejudice and discrimination in Britain. To know ‘what works’, it is essential that those organisations that are working hard to carry out anti-prejudice interventions are able to evaluate what they’re doing sufficiently well that others can be confident in the information and insights produced by project evaluations.**

In July 2016, the Commission published its research report ‘Prejudice and unlawful behaviour: exploring levers for change’ by Abrams *et al.* This publication set out the evidence in Britain on how prejudice relates to unlawful discrimination, and identity-based violence and harassment. The report identified a need for better evaluation of which anti-prejudice interventions are effective and why. In response to this, the Commission asked us (the Coalition for Racial Equality and Rights) to develop guidance and training on how to evaluate interventions for organisations working to tackle prejudice, discrimination and identity-based violence and harassment.

We were asked to identify a minimum standard of evaluation that will allow organisations to evaluate in a way that is proportionate and realistic. This means taking into account the barriers to evaluation, but also ensuring the process is sufficiently robust and transparent to allow the Commission and others to be confident in the findings and use them in designing policy.

To produce this guidance:

- **We reviewed published literature on effective approaches to the evaluation of attitude and behaviour change interventions.**
- **We reviewed published guidance and toolkits on evaluation.**
- **We interviewed evaluators, practitioners and managers about their experience of evaluating relevant projects.**
- **We tested elements of the guidance with organisations delivering anti-prejudice interventions through capacity building and feedback on the capacity-building sessions held in England, Scotland and Wales (see Appendix for more detail).**



# The value of evaluation

The core benefits of evaluation are the potential to improve the effectiveness of your work, to make a greater difference to prejudice, and to get more impact from the time and resources you deploy on any intervention. However, evaluation itself takes time and resources. If your organisation is to invest time and resources in evaluation, there should be a benefit to the organisation. Evaluation can help produce the following organisational benefits:

- **You have better evidence of how effective your work is.**
- **You have an improved understanding of how your work makes a difference.**
- **Your work is more effective and more cost effective.**
- **Your organisational planning and strategy are more focused on what works.**
- **The value of your work is more persuasively demonstrated to funders, partners, service users and other stakeholders.**
- **You are better able to attract new or increased funding.**
- **Your staff and volunteers feel a greater sense of achievement.**
- **You are better able to influence policy in your area of work.**

Being able to specify the benefits to your organisation can help make the case for greater organisational investment in evaluation, and can ensure that the reasons you are spending time and resources on evaluation remain clear throughout the whole process.

The idea that organisations need an incentive to conduct ‘good enough’ evaluations may sound a little counter-intuitive. Surely any organisation concerned with tackling prejudice wants to know that its work is effective, and how to make it more effective? In truth, although few organisations would disagree, there are also some powerful organisational disincentives for committing to candid, objective, ‘warts and all’ evaluation.

Anti-prejudice interventions are often delivered as projects funded by a third party. That means that the design and implementation of the project are conditioned by a consideration of the funder’s requirements and intended outcomes. Funders normally require reports on the projects they fund. Consequently, evaluation can be seen as an activity to satisfy the needs of funders.

This can mean that a desire to present ‘good news’ to a funder, and to attract further funding, produces a competing agenda that can skew evaluation processes and findings.

## Learning from a programme-level intervention

In 2012 the Scottish Government introduced funding to support a number of pilot projects, intended to produce learning on what works in tackling sectarianism. The Voluntary Action Fund was commissioned to manage the fund and to support funded projects to produce useful evaluation reports. Despite considerable investment from the Voluntary Action Fund in promoting a message that it was acceptable for projects to report limited success, as long as there was useful learning in it, funded projects still focused heavily on ‘success stories’. Where learning was identified, it tended to focus on learning about the operational aspects of project implementation rather than the fundamental assumptions underpinning intervention design, which may have produced more useful learning.



# Barriers to evaluation

**Despite the clear benefits of conducting ‘good enough’ evaluation, the evidence tells us that evaluation is not always done, and not always done well.**

There are numerous reasons for this, including:

- **Resourcing**
  - » Evaluation can be costly and time consuming.
  - » Organisations prioritise resources for service delivery.
  - » When planning evaluation, organisations fail to adequately specify and commit the resources required.
- **Evaluation design**
  - » The scope of the proposed evaluation is unclear or contested.
  - » There is no baseline from which to measure change.
  - » The things that organisations evaluate are not the outcomes of their work.
- **Skills for evaluation**
  - » Organisations are not sure how to evaluate the outcomes of their work.
  - » Organisations don't know how to turn the data they collect into conclusions.
  - » Evaluation isn't used to produce learning or inform change.
- **Conflicting motivations**
  - » Organisations are reluctant to evaluate in case they learn things they don't want to expose.
  - » Organisations' main motivation for evaluation is to satisfy funders or other stakeholders.
- **External influences**
  - » Short-term project funding makes evaluation after the end of a project difficult to resource.
  - » Evaluation is considered too late in project planning and delivery processes to be designed effectively.

Evaluation can be costly, and it can often seem like doing the project or intervention itself is more important. However, consider the potential costs of doing work that's ineffective, or that could be more effective, for years. How costly does evaluation seem now?

Like any business proposal, the business case for evaluation must be persuasive. The benefits delivered by evaluation need to be commensurate with the resource input required. An evaluation design should be proportionate to the scale of the intervention and the return it promises.

Designing and delivering an evaluation is a skilled and complex task. Like many complex tasks, evaluation can become more manageable when broken down into its constituent parts. Evaluation design and delivery is a matter of establishing clarity on what you hope to achieve through evaluation, how you will go about it, how you will resource the process, and how you will use the findings.

The aims and scope of externally commissioned evaluations are normally specified in an evaluator's brief. Methodology and resourcing are then agreed and recorded in a work programme produced by the evaluator and agreed by the client, and project management processes are put in place. When undertaking self-evaluation, organisations rarely apply this level of attention or project management.

This guidance is designed to support you to produce a plan for self-evaluation that is equivalent to an internal commissioning process. An internal commissioning process can act as both a business case for evaluation, helping to persuade others in the organisation why evaluation is worthwhile, and as a project plan for evaluation. An internal commissioning statement should set out:

- **the aims and scope of the evaluation**
- **the baseline from which change may be measured**
- **the activities, outputs and outcomes of the intervention being evaluated**
- **the success measures required to show progress on outcomes**
- **the methods of collecting data on success measures**
- **the processes for turning that data into conclusions and learning**
- **the timescales for each element of the evaluation**
- **the responsibility for leading on each element**
- **the resources, including staff time, required to deliver the evaluation**
- **how the final report from the evaluation will be used.**

Evaluation can expose inconvenient truths. It can show that work in which you have invested time, energy and resources has not made any real change. It can show up flaws in intervention designs that you may have been wedded to for long periods. It can demand that you think and act more rigorously about what you are doing, and why you are doing it in the way that you are. This can be uncomfortable, but should be beneficial

in the long term. Your work will become more effective, will be seen to be more effective, and will be better placed to attract greater support from others.

'Measuring outcomes' guidance produced as part of a US capability-building project for non-profit organisations (Compassion Capital Fund National Resource Center, 2010), notes:

**'Although there are many uses for the information generated by outcome measurement, organizations often make the effort because they are required to do so. They are asked to be accountable for the use of their grant maker's funds.'**

When undertaking evaluation for external stakeholders, it is tempting to present your project in the most positive possible light, especially if they have some power over your future funding or status, or the wider perception of your organisation.

Although this may not seem unreasonable, by doing so you are in effect distorting the truth of the evaluation and limiting the potential of the evaluation to really make a difference to prejudice, discrimination, and identity-based violence and harassment in Britain.

One of the most frequently expressed barriers to evaluation is a perceived conflict between the reporting demands of funders and the kind of evaluation that projects can use to establish the value of their approach and lessons to inform future work. There really is no reason why this should be so. Funders have their own outcomes that they wish to progress by funding other organisations. It is quite reasonable and appropriate for funders to seek evidence of how their investment has helped produce the outcomes that they are interested in, and for them to ask funded projects for this evidence.

It is equally reasonable and appropriate for funded projects to have their own planned outcomes and priorities. There may be situations where funders' reporting requirements and the project's ideal evaluation processes overlap to different degrees, but one does not necessarily prevent the other.

When dealing with funders, projects may be relatively powerless to influence what the funder requires of them, and attempts at negotiation risk producing frustration. A less frustrating approach may be to say to funders 'Yes, and ...' instead of 'Yes, but ...' In other words, projects may comply with their funders' reporting requirements and undertake evaluation processes that meet their own needs.

# A minimum standard for evaluation

**The Commission's report 'Prejudice and unlawful behaviour: exploring levers for change' by Abrams *et al.* (2016), applied a set of criteria as a method of assessing and scoring the quality of the evaluation processes of 24 different interventions.**

The scores achieved by the 24 interventions studied ranged from 15 per cent to 71 per cent. The variation in the quality of evaluation processes that were reviewed show just how challenging it can be to undertake a 'good enough' evaluation.

We have designed these principles following analysis of the barriers and challenges to enable you to achieve that 'good enough' evaluation. By 'good enough' we mean an evaluation that is proportionate and realistic, but can also provide evidence that is reliable enough for the Commission and others to use to inform their work.

It is intended to be used as a planning tool, so consideration of resourcing is placed near the end, when you will have sufficient clarity on the evaluation work programme to make an informed assessment of the resources required to deliver it. The principles can also be used to assess evaluations while they are still going on, to establish which elements are good enough and where there are opportunities for improvement.

By applying these principles when planning a new intervention, you will be better able to establish if, and how, that intervention made a difference. If you are already delivering an intervention, you can use the principles to interrogate your processes and identify opportunities to improve your evaluation.

**Figure 1**

Eight principles for evaluating anti-prejudice projects



# Designing the project

## Principle 1: Our decision to make an intervention is based on a robust assessment and specification of the need to make an intervention

It seems colossally obvious to say that, before embarking on an intervention, you should be sure that there is a need for it. However, many interventions are founded on vague, woolly or out-of-date ideas of need, or an unfounded perception of need created by media or political pressure.

Your understanding of the need for an intervention should be clear, specific, evidence based and up to date. Ask yourself:

- **What is the nature of the problem?**
- **What evidence is there that it is a problem?**
- **What do you know about the scale or extent of the problem?**

You may find that there is insufficient evidence to support any kind of intervention. Establishing evidence of need can prevent us from investing resources in unnecessary interventions.

Asking yourself these questions can clarify how the problem is manifested. For example, it might seem apparent that there is a stigma associated with having a mental health condition. Now ask yourself, what evidence do you have that there is a stigma associated with mental health? What are the factors that can be observed and counted that mean you can confidently say that there is a stigma and how it manifests itself? Understanding how a problem manifests itself can help with considerations of how to tackle that problem.

Knowing how the problem manifests itself can help produce a baseline from which you can measure the change delivered by your intervention. The baseline is the foundation of any evaluation. If you don't know where you started from, you can't tell how far you've come.

There are two forms of evidence that can help to establish whether there is a need for an intervention. Secondary evidence is data that has been collected by other parties. Normally it will have been analysed and published in the form of a report that draws



out and presents the findings in an accessible format. This should already be quality assured, so you can be sure of its validity, and should be relatively easy to obtain.

However, the specific data that you need to inform your project may not be available from secondary sources, particularly if you are working at neighbourhood level or dealing with manifestations of prejudice where there has been no published data.

Large-scale social attitudes surveys such as the British Social Attitudes Survey and Scottish Social Attitudes Survey can be useful. The fact that these sources of data are published annually can enable you to establish trends; how attitudes and behaviours have changed over time.

Sometimes you will exhaust the sources of secondary evidence and still not be confident that you could convince a doubter of the problem. This doesn't mean that there is no problem. It means that the available evidence is insufficient to form conclusions on the level or nature of the problem. At this point, it may be necessary to conduct some primary research to clarify the extent of the problem.

Primary evidence is data that you obtain by going direct to the source and conducting your own research. This can produce more detailed, nuanced and relevant information. However, it can be time consuming, and (unless you have the skills and resources to do it robustly) can produce data of questionable quality.

Undertaking primary research can be quite simple, or very complex, depending on what you aim to explore. If data exists at national or regional level, but not at the geographic level you need, you may be able to mirror the proven methodology and the question design used in the available research. This will give credibility to your research and allow you to easily compare your findings to the national or regional data.

When dealing with prejudice, there can be additional sensitivities associated with simply asking people about their experiences as victims or perpetrators. Almost all social research processes are subject to what researchers call 'response bias'. Response bias takes place when people answer questions in a misleading or untruthful way, and can occur both consciously and unconsciously. This is a particular risk for evaluating anti-prejudice interventions, as there are many different motivations for bias in this context.

The way that questions are asked can also influence responses. 'Prejudice and unlawful behaviour: exploring levers for change' by Abrams *et al.* (2016), notes that two-part questions asking respondents firstly if they have been discriminated against, and secondly why they believe the discrimination took place, generate lower estimates of prevalence than if respondents are asked in the first instance about their experiences in relation to a protected characteristic.

A common reason for needing to undertake primary research is because the target population you are considering is relatively specific, and the data is simply unavailable

at that level. There may be specific populations who are demonstrably more likely to have prejudiced attitudes, or who are more at risk of prejudice-based discrimination than others. Good-quality evidence can mean that your intervention can be targeted at those who will most benefit from it, and can be designed using the available evidence on what works with those target groups.

Sometimes there can be ethical or practical considerations that prevent interventions being targeted only at a particular group of people, and as a default the intervention is targeted at the wider population. In situations like this, for example in national attitude and behaviour change campaigns using marketing methods, it can be relatively easy to establish change using national data, but very difficult to confidently attribute that change to the intervention.

### Intervention without evidence

In 2009, one of the authors of this report worked with an organisation in a London borough that had spent two years on a project designed to overcome tensions between young people of different faiths following the 7/7 London bombings. Other than the statements of practitioners, the organisation had little evidence of tensions between young people of different faiths before or after the bombings, and no intelligence on young people more likely to be involved in conflict.

Unsurprisingly, the organisation was unable to evaluate if this project had made a difference to tensions between young people of different faiths, because it had no robust evidence that such tensions existed before its work.



Many of the techniques to collect data for evaluation purposes described in this guidance may easily be applied to conducting primary research. By using the tips in the guidance to design a methodology for conducting primary research to establish the extent of a problem, you should be able to produce a reasonable baseline from which to measure any change arising from your activity.

## Principle 2: We have clarity about the difference we wish to make

By applying Principle 1, you should have a clear idea of the nature of the need for change. If you are vague about the nature of the change you want, then you will find it impossible to measure if you have achieved it. The next step is to set out your vision of the nature of the change that's needed in an outcome statement. This represents outcomes in a 'theory of change' approach.

Outcomes and outputs are frequently confused. Charitable think tank New Philanthropy Capital define outputs as 'products, services or facilities that result from an organisation's or project's activities' and outcomes as 'the changes, benefits, learning or other effects that result from what the project or organisation makes, offers or provides' (Kazimirski *et al.*, 2016).

Outcome statements are essentially a description of the desired change. For example, 'organisations delivering anti-prejudice interventions have increased capacity to evaluate the change they make' is a statement of outcome.

For the purposes of evaluation, outcomes should be specific enough to be measurable, and for change to be reasonably attributable to the intervention. Broader outcomes of societal change and 'world peace' tend to be unsuitable for project design, or for evaluation at project level. They might be worthwhile having in a statement as a reminder of our long-term aspiration, but won't be achievable by, or attributable to, a single intervention.

Outcomes for anti-prejudice interventions should, as far as possible, specify the protected characteristic that they intend to affect. It can be tempting to focus on less specific ideas of equality and prejudice, but change in these concepts can be more difficult to demonstrate than change in a specific circumstance associated with a specific protected characteristic. While these broader changes in values or attitudes may be desirable, it is exceptionally difficult to generate baseline data on existing levels of non-specific prejudice, and therefore difficult to confidently measure change.

If you want to have a manageable evaluation process, it is wise to choose fewer outcomes. This may mean prioritising the outcomes that you wish to invest in evaluating. The New Philanthropy Capital 'four pillar' approach to planning evaluation (Kazimirski and Pritchard, 2014) recommends that you should prioritise the most important outcomes and focus on measuring those.

Outcomes should also be expressed in such a way that they can be used for long periods of time, and possibly across different interventions, without amendment. This means that specific targets for change are not outcomes. For example, 'there is a 20 per cent reduction in reported incidents of discrimination' is a target, not an outcome, because it may be necessary to change the target reduction at some point.

## Principle 3: We have reason to believe that the intervention we propose to deliver will produce that difference

Once you have evidence that there is a clear and specific need, and you can clearly express the difference you want to make, you can move on to consideration of what you are going to do to produce that difference. You need to be able to show that you have a reasonable basis to believe that the activity you propose will produce the change you seek.

Very often interventions are designed on the basis of the commonly quoted ‘politician’s syllogism’: ‘Something must be done. This is something. Therefore, we must do it.’ Just as frequently, interventions are designed on the basis of ‘common sense’; people do what seems sensible. The risk with such interventions is that they are not as effective as they might be, or can in fact have a negative impact. As far as possible, interventions should be selected and designed on the basis of evidence that the activity proposed actually works.

The Commission’s report ‘Prejudice and unlawful behaviour: exploring levers for change’ by Abrams *et al.* (2016), states that: ‘Well-designed interventions should be based on a review of the literature.’ Interventions that have been studied and proven to work are better than interventions that have not been proven. However, the available literature on effective anti-prejudice interventions is limited, and some conclusions may be contested.

In some cases, because available evidence may be limited or questionable, it may not be possible to design an anti-prejudice intervention from the literature alone. In other cases, you may want to test out innovative ideas that build on the lessons from the literature. The potential to add to the pool of evidence that is available to inform other interventions is a further reason why ‘good enough’ evaluation is important.

If you choose to undertake activity that has not yet been shown to be effective, you need to be able to show why you think it might be effective. A ‘theory of change’ sets out what a project intends to achieve, and how that will be achieved. Nesta, the UK-based innovation foundation, offers a particular model for developing a theory of change that may be useful for anti-prejudice work (Nesta, 2011a; Nesta, 2011b). The model provides a simple visual diagram that prompts you to ask and answer a set of simple questions to clarify the process through which your activity will create change.

The theory of change is useful because it asks you to define the problem being tackled, as this guidance does in Principle 1. Some other planning tools skip this step, beginning instead with the change you want to achieve, as this guidance does in Principle 2. Starting at Principle 2 can mean that your understanding of the problem itself is flawed or incomplete, making it harder to both plan and evaluate the project.

A theory of change is a good starting point, but it is designed as a planning tool, not an evaluation tool. Theories of change ought to be statements of our assumptions (Nesta, 2016), but they can easily become regarded as statements of ‘what is’. A theory of change should act as your hypothesis, a statement of what you believe is reasonable but plan to test using evidence from evaluation.

### Counter-intuitive interventions

Since 1992 the Changing Faces charity has been working to help people who have a disfigurement find a way to live the life they want. In that period, it has learned that ‘common sense’ and good intentions are not sufficient to be confident that an intervention is positive. When a child with a disfiguring birthmark, condition or injury starts school, staring, teasing and rejection are real possibilities (Rubin and Wilkinson, 1995). Preparations by schools therefore often include asking children not to stare, and giving a talk about whatever it is that makes their new classmate look unusual. However, making friends begins with looking at each other (Bronfenbrenner, 1979). Prohibiting staring increases isolation (Frances, 2004) and may be behind the finding that these children tend to have fewer school friends than children whose appearance is not unusual (Kish and Lansdown, 2000).

Giving children information about a child’s condition, injury or treatment also makes things worse (Vandell *et al.*, 1982). It seems to send the unintended message that this child is not like them. It damages children’s main way of getting to know someone new – asking questions. Children’s attitudes to difference are much more positive following opportunities to debate about disfigurement, disability and what constitutes bullying (Casalme, 2016).

If an intervention is not evidence based, its evaluation may be very disheartening.



Marcus, who has a cleft lip and palette, has now moved to secondary school along with friends Marcus Brookes, left, and Connor Pugh, right.

*Photograph: Christopher Thomond for the Guardian*

# Planning the Evaluation

## Principle 4: We have clarity on the nature of the data required to demonstrate that we have made a difference

The guidance in Principles 1–3 helps us to make good decisions on what we want to achieve. The key question for evaluation, however, is ‘how will we know if we have achieved our outcomes?’ Outcomes are often intangible, so it’s essential to be able to specify more tangible things that allow you to make an informed assessment of the extent to which your activities have delivered progress towards your outcomes.

Imagine your planned outcome is ‘young people have increased confidence to challenge their peers’ use of prejudiced language’. Confidence is intangible. You cannot see or touch or count the change in an individual’s confidence. You need to specify things that are observable that can allow you to reasonably conclude that a change in confidence has taken place. These observable elements can be called ‘success measures’.

Examples of success measures that could achieve this include:

- **number of young people reporting peers’ use of prejudiced language to school staff**
- **number of observations of young people avoiding contact with peers who use prejudiced language**
- **number of young people reporting that their confidence to challenge their peers’ use of prejudiced language has increased.**

Normally you will specify two or three success measures for one outcome. This is because a single measure on its own may not be reliable enough to allow you to reasonably conclude that the outcome has been met. This is particularly true of success measures in which participants report their own perceptions of themselves or their behaviour. Even when they’re being honest, people are not always the best judges of their own characteristics, and these sorts of responses can be influenced by response bias and social desirability concerns. Further questions may be needed to make sure you have a more complete picture of whether the success measure has been met.

Remember that if you have the evidence of how the problem you are seeking to tackle manifests itself, you may already have the success measures that allow you to

demonstrate change. If you have based your intervention on published literature, the success measures may already be in the literature.

To be useful, success measures should pass the following tests.

**Clarity:** A success measure should be clearly enough expressed that its meaning cannot be open to different interpretations.

**Reliability:** A success measure should produce consistent results when used to measure change. If it doesn't, then there may be issues with ambiguity, and the findings of any evaluation may not be replicable.

**Relevance:** Inference is the act or process of deriving a logical consequence conclusion from premises. A success measure should tell us something about the change in the outcome to which it is applied – we should be able to infer that something has changed. It's quite common to find projects using success measures that don't actually allow them to logically infer anything about the intervention's planned outcomes.

### Irrelevant success measures

The outcome of the project described in Principle 1 was that tensions between young people of different faiths following the 7/7 London bombings had reduced. To achieve this, the project supported young people to produce and distribute a range of media messages promoting interfaith harmony.

The project's success measure was that 'young people involved in the production of messages enjoyed the opportunity and developed skills'. They could successfully show that the young people involved in the production of messages enjoyed the opportunity and developed skills. But because this measure of success cannot be logically linked to the planned outcome, it cannot be regarded as a relevant measure of success, and the evaluation cannot show change in the planned outcome.

**Measurability:** A success measure needs to be able to be counted in some numeric fashion. If it cannot be counted in some numeric fashion, you can't robustly show change from the baseline. This needn't be an absolute number; for example, a success measure can show change in frequency, proportion or range.

**Realism:** You should be able to collect the data that forms your measure of success. You may come up with success measures that would be perfect in an ideal world, but if you don't have the time to collect the data, or staff don't have the skills to observe and record it, then you won't be able to collect it.

## Challenging methods

In the Changing Faces example described in Principle 3, the measures of success in changing children's attitudes to their facially disfigured peers were the behaviours of the children. Skilled researchers spent many hours observing and recording the interactions between children in class and at play in a number of different locations, and many further hours analysing the frequency of specific behaviours. This method of data collection produces rich and robust results, but is challenging because it requires a high level of skill, objectivity and time.

**Ethics:** The process of collecting the data should not negatively affect participants. Some interventions, and some methods of measuring self-perceptions, may require that participants are not wholly informed of the intent of the process. Check with relevant ethical guidelines for your specialism or profession if there's a potential issue.

If there are no ethical guidelines governing research in your specialism or profession, the Social Research Association has developed guidelines (Social Research Association, 2003) that may be applied more generally.

**Burden:** This is really more of a concern for the method of data collection rather than success measure design, but it is worth mentioning here. If the collection of the data is so time consuming or burdensome for participants that it has a negative effect on participant experience or outcomes, then you should consider if the success measure is sufficiently important to warrant this.

As we have already noted, all evaluation should be proportionate. If we can objectively judge that the investment of time, energy and resources appears to be greater than the benefits available, it is quite reasonable to limit the scope of the evaluation accordingly. Evaluation is a balance between the ideal and the achievable, being the best you can achieve in the balance of cost, quality and time.

It is better to evaluate to the best of your ability than not evaluate at all. The limitations of your evaluation and the caveats you have to place on your findings can be detailed in your evaluation report. The New Philanthropy Capital publication 'Balancing act: a guide to proportionate evaluation' (Kazimirski *et al.*, 2016), offers further guidance on ensuring your investment in evaluation is proportionate to the scale of your intervention and the benefits the evaluation will deliver.



## Principle 5: We have clarity on the methods we will use to collect that data

One of the places where self-evaluation frequently goes awry is that organisations leap to thinking of methods of data collection without thinking about the earlier stages of intervention and evaluation design. Methods of data collection should largely be informed by the nature of the data to be collected, and the characteristics of the people who you may be collecting it from.

### Measuring long-term impact

A common method of evaluating the outcomes of anti-prejudice interventions, particularly those that take the form of short-term educational inputs, is to conduct pre- and post- testing. To measure a change in prejudice-based attitudes, for example, participants will be asked to respond to questions on their attitudes to people with a particular protected characteristic before and after the intervention. The difference between the attitudes at both points is therefore attributed to the activity. This form of data collection is useful because it is relatively simple to manage, produces a small volume of easily analysed data, and is comparatively free of confounding factors (variables that might make it difficult to be sure that it was your intervention that created the change).

This method does have drawbacks, primarily in that it doesn't measure how long change is sustained after the completion of the input. This can be addressed by returning to the target population at later dates, assuming that the resources are available and the systems are in place to retain contact with them.

### Capturing reliable information

More importantly, it is essential that pre and post testing is performed in a way that minimises the opportunities for the data collected to be challenged. When evaluating prejudiced attitudes, it's essential that questions are worded and administered appropriately. Under Principle 1, this guidance has already discussed the role of response bias, social desirability concerns and question design in undermining the reliability of data. Some things you may wish to think about when designing your success measures include the nature of prejudice experienced by the specific group of people you're working with. This may differ across different groups sharing protected characteristics, with different underlying attitudes and different manifestations. As well

as asking a number of questions to understand progress towards a particular success measure, it's important to develop a number of success measures that can demonstrate change in different types of attitudes and/or behaviours towards the intervention's overall aim (Peer and Gamliel, 2011).

## Measuring explicit and implicit attitudes

When evaluating an intervention designed to tackle prejudice, you may want to look at its effect on explicit or implicit attitudes. Harvard University's Project Implicit (2011) defines an explicit attitude as 'the kind of attitude that you deliberately think about and report'. It says that 'implicit attitudes are positive and negative evaluations that occur outside of our conscious awareness and control'.

Implicit and explicit forms of prejudice can't be measured in the same way, because explicit prejudice measurement relies on the subject being aware of their own attitudes or behaviours. To measure explicit prejudice, methods openly asking about a person's perceptions, attitudes or behaviours can be used. To evaluate implicit prejudice, however, a different approach is needed, which draws out underlying attitudes that the person holding them may not themselves be consciously aware of. This could include asking someone to 'fill in the blanks' in an unclear scenario or to describe an association with an image or object. A common method is known as the 'implicit association test' or IAT (Greenwald *et al.*, 1998).

### Implicit association testing in use

Scope is a national charity that exists to drive change across society so that disabled people have the same opportunities as everyone else. End The Awkward is a project of Scope's that aims to influence people's attitudes towards disability and reduce the awkwardness that can characterise encounters between disabled and nondisabled people.

Scope was aware that it can take time for attitudes and behaviours to change. Measuring change over the period of a time-limited campaign therefore could be challenging. To manage this, Scope used implicit association testing to test subconscious associations and ideas around 'disability'. By measuring how quickly people responded – and therefore how long they were thinking about their answer – Scope was able to successfully detect smaller, subconscious changes in attitudes over a six-week period.



## **Attitudes and behaviour**

Attitude change and behaviour change need to be measured with separate success measures, as they are separate factors. Attitudinal change can be measured using the methods described above. Behavioural change requires us to describe and observe actual behaviours.

If you have used secondary data to establish the prevalence of discriminatory behaviours, you may be able to return to further iterations of the same data sources to establish if there has been a change during the period of your intervention. If you have used primary sources, you may have to repeat the processes again to establish if things have changed. You should consider whether this could be time consuming and burdensome for your participants.

## **Sampling: choosing who to include in the evaluation**

Unless your intervention works with a small number of people, it may not be possible or reasonable to attempt to involve them all in its evaluation. If it works with more people, you need to select a sample to involve. 'A sample is a proportion or subset of a larger group called a population ... A good sample is a miniature version of the population of which it is a part – just like it, only smaller' (Fink, 2003). If the sample of participants used in your evaluation is not broadly representative of your total population, then your findings should be caveated to reflect the difference. Where you are working with small populations, you should take care that no individual may be identified from any report made public. If that should be the case, there are likely to be both ethical and data protection implications, and it may not be possible to share the report from your evaluation with people outside your organisation.

## **Taking a qualitative approach**

Much of the above may seem scientific and clinical. You may have experience of using more creative and expressive methods to enable participants to say, in a more open fashion, what an intervention has meant for them. Using methods like depth interviews, focus groups, video diaries, talking walls and illustration produces qualitative data, which can give you a greater sense of understanding the experiences, perspectives and perceptions of the evaluation subjects. The Evaluation Center at the University of Western Michigan notes that 'qualitative methods are often used in evaluations because they tell the program's story by capturing and communicating the participants' stories' (Patton, 2003).

Qualitative methods on their own are rarely suitable for the purposes of evaluation, but can be valuable complements to more quantitative measurements of change. They are particularly useful in the later stages of evaluations to assist in establishing how or why an effect has taken place. For example, if your quantitative evaluation has shown that, following young people's exposure to positive images of older people, there is a reduction in their negative perceptions of older people, you could use focus group discussion to explore what it was about the intervention that produced that effect.

Qualitative data is more open to differing interpretations, and therefore offers a greater risk of inaccurate interpretation, or of the evaluation being perceived to be subjective or biased. It also means that qualitative data can fail one of the key tests for a good success measure – that it should be reliably and consistently interpreted by different parties.

The Evaluation Center's 'Qualitative evaluation checklist' (Patton, 2003) further notes that the 'quality of qualitative data depends to a great extent on the methodological skill, sensitivity, and integrity of the evaluator ... Generating useful and credible qualitative findings through observation, interviewing, and content analysis requires discipline, knowledge, training, practice, creativity, and hard work.'

In this respect, the data produced through qualitative methods may be regarded as potentially persuasive and useful in influencing others, particularly at reporting stage, but they present substantial challenges in producing data that is able to provide a robust evaluation of change from a baseline for your success measure.

# Carrying out the evaluation

## Principle 6: We know how we will analyse the data we collect to produce conclusions

Another area where self-evaluation processes frequently stall is in the analysis of the data collected. Organisations can quite competently design interventions, set outcomes and success measures, and use appropriate methods to collect the data, and then do very little with it.

Analysis can take time, but the key to efficient and effective data analysis is making the process manageable. That's achieved by only looking at data that allows you to form conclusions. It's not uncommon to invest time in gathering data you can't or won't use in your evaluation.

To prevent this, you can ask yourself a few simple questions:

- **'What will I do with this information?' If the answer is 'nothing' or 'I don't know', you may be better not collecting it.**
- **'How will I analyse the data I collect?' If the answer is 'I don't know' or 'with great difficulty', then you are either collecting data that doesn't meet the relevance test for success measures, or data that is qualitative in nature.**

Data that allows you to form conclusions about the change from your baseline is what you need for evaluation; that means data that allows you to measure difference, and that means quantitative data.

There can be a temptation in anti-prejudice work to prefer qualitative data. Qualitative data may allow you to describe circumstances or change in a narrative format, but it does not allow you to measure the difference between two sets of circumstances. Its use in evaluation is therefore limited to illuminating points you may wish to make in a more persuasive manner. Qualitative data may be used in reporting, but not in evaluating. It may therefore be more appropriate to apply qualitative methods after you have completed an evaluation and know what your successes have been, and therefore can make informed decisions about the qualitative data to collect and to include in reports and other evaluation outputs.

There is an exception to this rule, but it involves turning qualitative data into quantitative data. You can do this with transcripts of interviews, focus groups or video diaries by setting codes for the concepts that are important to your evaluation, and then going through the qualitative data to count how frequently each concept is raised. It can get a little more complex than that. You may want to note something of the context in which each concept is raised. This means that you need to specify the types of contexts and count them too. Researchers call this process 'multivariate data analysis', because it involves considering how more than one variable affects an outcome. This sort of analysis is often better approached by a professional researcher.

Using quantitative data means that your analysis can be relatively simple and, if you have the ability to use freely available numerical software packages, it can take very little time. If you set your baseline properly before embarking on your evaluation, then you should be able to easily decide what you need to know about the difference from that baseline.

There are some key quantitative data analysis concepts that can allow you to show what your intervention has achieved:

- **Number:** How many times in a set period can a success measure be observed? For example, the total number of reported crimes against older people in 2012.
- **Frequency:** How often does it take place in that period? For example, a crime against an older person took place every three days in 2012.
- **Distribution:** The total number of success measures arranged in order, for example, all crimes against older people in 2012 by date. This can be used to establish patterns, such as the months in which more of these crimes occur.
- **Range:** The difference between the lowest and highest values. For example, there were 100 crimes against older people in one location and 60 in another.
- **Mean:** Commonly called the 'average', although in technical terms it is one of several different types of average. The total sum of all values, divided by the total number of all values; for example, the average age of victims was 65.
- **Median:** The middle number of a range; for example, the median age of victims was 60.
- **Mode:** The most frequently occurring factor in a range; for example, the majority of victims were in the 70–75 age band.
- **Outlier:** A piece of data that lies far outside the normal range; for example, one victim was aged 110, but the majority were in the 60–80 age range. Outliers are often a result of errors in data collection. Even when they are not, they can skew other analyses of the data and are often excluded for this reason.

Data analysis does not have to be done at the conclusion of an intervention. Data analysis can be done at any significant point, either at the end of a set period of time, or once particular milestones have been reached. Time should only be invested in data analysis when there is a clear opportunity to use the results, either to improve the intervention as it is being delivered, or to make a judgement on its effectiveness once it is concluded.

## Principle 7: We know how we will use our conclusions

There is little point in investing in any evaluation process unless you plan to do something with what you learn from it. Before you begin to think about using your conclusions, you should know how confident you are in them. The greater your confidence in your conclusions, the greater the credibility your evaluation should have, and the more likely it is to influence change.

Nesta's 'standards of evidence' (Puttick and Ludlow, 2013) – level 1 being the lowest standard and level 5 the highest – were developed to allow users to assess the evidence behind programmes, products and services. Seeing where they are placed on the standards of evidence enables users to understand how confident they can be in their claims. The standards are as follows:

- **Level 1:** You can give an account of impact. By this we mean providing a logical reason, or set of reasons, for why your intervention could have an impact and why that would be an improvement on the current situation.
- **Level 2:** You are gathering data that shows some change among those receiving or using your intervention.
- **Level 3:** You can demonstrate that your intervention is causing the impact by showing less impact among those who don't receive the product or service.
- **Level 4:** You are able to explain why and how your intervention is having the impact you have observed and evidenced so far. An independent evaluation validates the impact. In addition, the intervention can deliver impact at a reasonable cost, suggesting that it could be replicated by (potential) customers and purchased in multiple locations.

- **Level 5:** You can show that your intervention could be operated by someone else and somewhere else, and scaled up, while continuing to have a positive and direct impact on the outcome, and while remaining a financially viable proposition.

By applying the principles in this guidance, you should be able to comfortably achieve levels 1 and 2 on the Nesta scale (Puttick and Ludlow, 2013). You may be able to partially achieve level 3, but the demand for a comparator or control group may prove challenging for the evaluation processes described in this guidance.

Level 4 is concerned with independent evaluation and an economic analysis of the costs of the intervention. Such analysis is excluded from this guidance, which is concerned with demonstrating the difference that an intervention produces.

Level 5 is concerned with demonstrating replicability. Achieving level 5 requires multiple evaluations of a similar intervention in different settings. You may be able to achieve level 5 over a period of years, particularly if you can show that your intervention has been replicated by other parties that are not associated with your organisation.

You should have set out your aims for the evaluation as part of the process of establishing a business case for investing in evaluation. If you want your evaluation to produce any change, you need to understand the interests and motivations of the audiences for your evaluation.

To decide what level of evidence you will aim for, ask:

- **What should be different as a result of what you know through your evaluation?**
- **Who should be doing something differently?**
- **Why should they be doing that? What might motivate or prevent them from doing so?**
- **What role does your evaluation have in influencing change?**

The answers to the above questions should help you frame the recommendations arising from your evaluation. Sometimes evaluations stop at a point where they have made a judgement about the value or effect of the intervention studied. This is perfectly acceptable, but if you really want your evaluation to influence change, it's a good idea to specify the change and name the parties that you think have the power to make that change happen. A simple stakeholder analysis can help you think through why particular parties may be interested in your evaluation.



**Table 1**

Example of stakeholder analysis

Who are they?	What do they want?
Managers in our organisation	<ul style="list-style-type: none"> <li>• Evidence to inform how they direct staff time</li> <li>• Evidence to inform how they allocate funds and prioritise services</li> </ul>
Marketing and communications department in our organisation	<ul style="list-style-type: none"> <li>• Good news stories</li> <li>• Awareness of potential bad news stories so that they can manage the associated risk</li> </ul>
Funders and commissioners	<ul style="list-style-type: none"> <li>• Evidence to justify their investment</li> <li>• Evidence that the project is being efficiently delivered</li> </ul>
Academics and researchers	<ul style="list-style-type: none"> <li>• Evidence on factors affecting impact</li> <li>• Specification and testing of models of intervention</li> <li>• Opportunity to publish research</li> </ul>

There are many tools for thinking about your stakeholders. One of the most commonly used is Mendelow's 'power-interest grid' (Mendelow, 1991), which allows users to plot stakeholder positions according to their power to influence a project and their interest in its processes and results. The primary purpose of this tool is to allow managers to make decisions about where to invest time on managing stakeholders. More recently, Lynda Bourne described a process for stakeholder mapping and relationship management known as the 'stakeholder circle' (Bourne, 2015).

## Evaluation outputs

If your evaluation is to influence change, then the people who need to lead on producing that change need to be exposed to your evaluation's findings. You need evaluation outputs that are accessible, credible and persuasive for your target audiences. This will almost always mean an evaluation report.

An evaluation report should set out:

- **the background to the intervention and the evaluation**
- **the nature of the intervention, including the evidence of need, and the rationale for the intervention design**
- **the aims of the evaluation**
- **the methodology for the evaluation, including data collection methods, the nature and scale of the data gathered, and the processes of analysis applied**
- **the findings of the evaluation**
- **the implications of the findings, including recommendations for change where possible.**

The Equality and Human Rights Commission's report 'Prejudice and unlawful behaviour: exploring levers for change' by Abrams *et al.* (2016) used a set of eight criteria to analyse the methodologies described in a set of reports from evaluations of anti-prejudice projects. These criteria were:

- **Design** – information on the strategy chosen to carry out the evaluation
- **Sample** – information on the people or units of data being assessed
- **Reproducible** – information that tells you how to reproduce the evaluation (or intervention)
- **Ethics** – information on particular ethical issues encountered, or ethical approval sought
- **Outcomes** – information on the types of measures that are assessed, and the strength of these
- **Findings** – information on how much these meet the aims, and whether they are generalisable or transferable, relevant to policy or practice, and contribute to your current knowledge
- **Analysis** – information on how measures were analysed, and how much the analysis used is appropriate to the aims and the design and measures used
- **Limitations and other follow-up** – information on other aspects of the evaluation, including any limitations, any assessment of cost effectiveness, and participant satisfaction.

The above may be useful as a quality assurance checklist for your evaluation report. Although your report may not wholly address all the above criteria – none of the reports assessed in ‘Prejudice and unlawful behaviour: exploring levers for change’ by Abrams *et al.* (2016), scored above 71 per cent – applying them will enable you to make an informed assessment of your confidence in the report, and to identify opportunities to improve future evaluation activity.

It can be a good idea to produce a range of different reports for different audiences. You may wish to produce a full technical report for other professionals with a significant interest in your work, and a short summary report for those with a more marginal interest. You may also wish to extract particular elements and highlight them in reports that are tailored to the interests of particular stakeholders.

Similarly, participants may not want to read a lengthy report in an academic or professional style, particularly if they face literacy or language barriers. You may wish to produce a report in a different style for them, or work with a small group to design a different sort of output for their peers.

Although we made it clear in Principle 6 that the process of evaluation is a matter of quantitative measurement, the process of using evaluation findings to influence change is more nuanced. Some decision-makers may be influenced by hard, numeric data, while others may be more influenced by the difference made to real people. If you have qualitative data that corresponds with, and highlights, your evaluation findings, you may use it effectively here. This is not technically evaluation, but is delivering on the aims of evaluation, and therefore should be considered in any process concerned with influencing change.

You will want people to be exposed to your conclusions. In the internet age, the challenges of publishing information are greatly reduced. However, the challenges of ensuring that your information stands out from all the other information available are greatly increased. You will need to do something to ensure that the right people are exposed to your conclusions and know why they should give these their attention.

This may mean publishing and sending out a report. It may involve meetings, seminars, conferences or other events. You may wish to involve other parties with expertise in communications at this point. If you planned your evaluation well at the outset, you will have already secured agreement from people whose input is needed.

## Principle 8: We have assessed and committed the resources required to deliver the evaluation

It may seem a little unusual to have this principle at the end. Surely we need to look at the resources we have available for evaluation first? Well, yes and no.

It's never ideal to design any project or evaluation around the resources you have. The ideal is to design the right project, and then to consider the resources you need to implement it. You may find that you don't have, or can't get, all the resources you need. At that point you may have to look again at your project design, but you will then be able to prioritise from an informed position, and reduce costs and activities that are inessential.

If you are commissioning an external evaluation, you will use a similar process of planning and negotiating. You will set out the rationale for the evaluation, what you need from it, and what you can put into it. The evaluator will look at the job that needs doing and specify a work programme to do it. Once they've done that, they will be in a position to estimate the time commitment needed, and consequently the costs to the client. If the client budget is insufficient, they can begin to look at revising the work programme to strike a balance between quality and cost.

When commissioning an external evaluation, the main financial outlay will be the cost of time from the evaluators you commission. When conducting an internal evaluation, you won't have those costs, but you will have a substantial time commitment from your own staff. Anti-prejudice interventions are often led by small project teams within larger organisations, or even a single person. You should designate a lead individual for the management of the evaluation and try to specify upfront how much time you will need from them, and from both internal and external colleagues.

You should try to consider:

- **Time for project management: Your evaluation, like anything else, needs to be kept on track and subject to quality assurance processes. Someone needs to do this, and they will need time to do it well.**
- **Time to design, test and refine data collection tools: Drafting questions for surveys, topics for interview and guides for focus groups takes a substantial amount of time. It's common practice to test and 'debug' these tools before using them more widely.**
- **Time to implement data collection processes: Interviews take time to arrange and undertake. Focus groups can also take a fair bit of time to arrange, and even online surveys can take quite a lot of time to promote.**

- **Time for data analysis:** Simple quantitative data analysis should not take long, if you have designed your questions well and do not have a large amount of data. You need to set aside time for this though, as an error in data analysis can make your whole evaluation dubious.
- **Time to write and edit the report:** This can be very time consuming, particularly if you are not used to writing evaluation reports.
- **Time for ‘Maxwellisation’:** Maxwellisation is the term used for the practice that allows people who are to be criticised in government reports to respond prior to publication, based on details of the criticism they receive in advance. If your evaluation, however implicitly, criticises any other party, they may take issue. It may be tactically wise to deal with this before publication. You can use a similar process to eliminate issues of factual accuracy, and ensure that you are forewarned of any challenges that may be made later.
- **Time for dissemination:** You will want people to be exposed to your conclusions; you will need to do something to ensure that they are. This may be as little as publishing and sending out a report. It may involve meetings, seminars, conferences or other events. Someone needs to commit the time to deliver these.

You may also need other resources to deliver your evaluation, including:

- **accommodation and refreshments for evaluation participants**
- **printing and/or design costs for surveys and reports**
- **incentive costs to encourage people to take part**
- **travel and subsistence**
- **dissemination costs, if you plan to hold any events to showcase your findings.**

When you are done, you can specify the resource requirement in your internal commissioning statement. This will help you influence colleagues, and ensure that they keep to their commitments when you go back to them some time after you first discussed evaluation with them.

# Conclusion

This guidance is intended to make evaluation as simple and achievable as possible, and to help organisations to evaluate their interventions in a way that is proportionate and realistic. Everyone reading this guidance is likely to be committed to the reduction or elimination of prejudice. If we don't evaluate well, we'll never know more about what works, and never be able to do better with the resources we have.

We know, however, that evaluation can become complex and hard to manage. You should not be put off by this. If you are still unsure how to go about evaluation, do what you can. Select one or two of the principles to improve and work on them. It's better to improve some of your processes than none at all. Once you feel you've accomplished those, you can look at building on your success by incorporating other elements of this guidance, or some of the other published guidance and support materials available.

What's important is that you can make evaluation work for you, for your organisation, for the people your activity is intended to benefit, and for the sum of knowledge about what works in tackling prejudice.

# Useful resources

## **Compassion Capital Fund National Resource Center (2010), 'Measuring outcomes'**

A comprehensive guide to techniques for measuring change in outcomes. Part of 'Strengthening non-profits: a capacity builder's resource library', created to continue capacity building work funded by the US Department of Health and Human Services.

## **Nesta (2011a), 'Guidance for developing a theory of change for your programme'**

A concise guidance paper on how to produce a theory of change and relate it to the Nesta standards of evidence.

## **Nesta (2011b), *Theory of change: define your goals and how you will achieve them.* [ONLINE]**

A theory of change worksheet to help you to map out the steps you need to take to achieve a particular goal, identify the potential impact and risks of your plan, and connect your work to a bigger goal.

## **Patton, M. Q. (2003), 'Qualitative evaluation checklist', University of Western Michigan**

A checklist to help you decide when qualitative evaluation methods are appropriate, what factors to consider and the most appropriate approaches. One of a suite of Evaluation Center checklists and resources.

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

Concise guidance on how to assess the quality of evidence produced by your evaluation.

## **Social Research Association (2003), 'Ethical guidelines'**

Accessible guidance for researchers on how to anticipate and manage ethical issues.

# References

- Abrams, D., Swift, H. J. and Mahmood, L. (2016), 'Prejudice and unlawful behaviour: exploring levers for change'**. London: Equality and Human Rights Commission. [accessed: 18 July 2017].
- Bourne, L. (2015),** *Making projects work*. Boca Raton, Florida: Auerbach Publications.
- Bronfenbrenner, U. (1979),** *The ecology of human development: experiments by nature and design*. Cambridge, Mass: Harvard University Press.
- Casalme, A.G. (2016), 'Engaging children in discussions of disfigurement and disability: the wonder of Palacio's Wonder',** *The Cutting Edge: The Stanford Undergraduate Journal of Education Research*, vol. 1, no. 1, pp. 26-44. [accessed: 18 July 2017].
- Fink, A. (2003),** *How to sample in surveys*. 2nd edn. Thousand Oaks, California: SAGE Publications, Inc.
- Frances, J. (2004),** *Educating children with facial disfigurement – creating inclusive school communities*. London: RoutledgeFalmer.
- Greenwald, A. G., McGhee, D. E. and Schwartz, J. L. K. (1998),** 'Measuring individual differences in implicit cognition: the implicit association test'. *Journal of Personality and Social Psychology*, vol. 74, no. 6, pp. 1464-80.
- Kazimirski, A. and Pritchard, D. (2014), 'Building your measurement framework, NPC's four pillar approach'**. London: New Philanthropy Capital. [accessed: 18 July 2017].
- Kazimirski, A., Harrison-Evans, P. and McLeod, R. (2016), 'Balancing act: a guide to proportionate evaluation'**. London: New Philanthropy Capital. [accessed: 18 July 2017].
- Kish, V. and Lansdown, R. (2000),** 'Meeting the psychosocial impact of facial disfigurement: developing a clinical service for children and families', *Clinical Child Psychology and Psychiatry*, vol. 5, no. 4, pp. 497-512.
- Mendelow, A. L. (1991),** 'Mendelow's power-interest grid'. Ohio: Kent State University.
- Nesta. 2016.** *What's wrong with theories of change?* [ONLINE] [accessed: 18 July 2017].
- Nesta (2011a), 'Guidance for developing a theory of change for your programme'**. [accessed: 18 July 2017].
- Nesta (2011b), Theory of change: define your goals and how you will achieve them.** [ONLINE] [accessed: 18 July 2017].



**Patton, M. Q. (2003)**, 'Qualitative evaluation checklist', University of Western Michigan. [accessed: 18 July 2017].

**Peer, E. and Gamliel, E. (2011)**, 'Too reliable to be true? Response bias as a potential source of inflation in paper-and-pencil questionnaire reliability', *Practical Assessment, Research and Evaluation Journal*, vol. 16, no. 9. [accessed: 18 July 2017].

**Project Implicit. [2011]**. Frequently asked questions. [accessed: 31 July 2017].

**Puttick, R. and Ludlow, J. (2013)**, 'Standards of evidence: an approach that balances the need for evidence with innovation'. Nesta. [accessed: 18 July 2017].

**Rubin, K.H. and Wilkinson, M. (1995)**, 'Peer rejection and social isolation in childhood: a conceptually inspired research agenda for children with craniofacial handicaps', in Eder, R. (ed) *Craniofacial anomalies: psychological perspectives*. New York: Springer, pp. 158-76.

**Social Research Association (2003)**, 'Ethical guidelines'. [accessed: 18 July 2017].

**Vandell, D.L., Anderson, L. D., Ehrhardt, G. and Wilson, K.S. (1982)**, 'Integrating hearing and deaf preschoolers: an attempt to enhance hearing children's interactions with deaf peers', *Child Development*, vol. 53, no. 5, pp. 1354-63.

# Appendix: Capacity building sessions in England, Scotland and Wales

**While developing this guidance to evaluation, the Coalition for Racial Equality and Rights (CRER) ran capacity building sessions using the guidance to increase the ability of organisations to evaluate anti-prejudice interventions.**

Three capacity building sessions on evaluating anti-prejudice interventions were delivered to 45 people in February 2017 in Glasgow (18 participants), London (21 participants) and Cardiff (6 participants). Participants were from a range of backgrounds, including civil society organisations, universities, public sector organisations and local authorities, all with some role that related to tackling or monitoring prejudice and discrimination. The training aimed to make participating organisations more confident in capturing the impact of interventions designed to tackle prejudice and discrimination, making the findings publicly available, and developing action plans for embedding evaluation in their practices and planning of future intervention work.

As the guidance was still in development while the capacity building sessions were being delivered, these provided an opportunity to use the information gathered through the sessions to inform the guidance. The focus of the training was on the core principles for evaluating anti-prejudice work, with interactive content designed to allow participants' views on barriers and solutions regarding evaluation practice to be gathered. The outcome of the exercises looking at barriers and solutions was used in developing the guidance, and is summarised here.

The final training content used in delivery consisted of a PowerPoint presentation with a series of accompanying handouts and exercise materials. The training was designed to be highly interactive, to cater for a wide range of learning styles and provide opportunities to gather learning for the guidance development work.

## Outcomes of the sessions

Participants were asked to complete an evaluation sheet at the end of the session. Most participants either agreed or strongly agreed that the information provided was useful (94 per cent), that the delivery methods were appropriate (88 per cent), that the training was accessible (94 per cent), and that they would recommend the training (85 per cent). Only two participants commented on methods, with one feeling that there should have been less group work and more input from the trainers, and another feeling that the training input was a little overwhelming and they would have preferred more time for reflection.

When asked how they planned to use their learning:

- **91 per cent said “share information with colleagues.”**
- **70 per cent said “build evaluation into our planning processes.”**
- **52 per cent said “evaluate a specific project (one you can already identify).”**

Participants used free text response boxes to describe how they planned to use their learning. Responses included:

- **“Consider what works and what is known not to work, especially when working on sensitive topics.”**
- **“Educate, inform, acquaint and recommend the practices and methods gained through the knowledge provided to my organisation”; and “Promoting an increased focus on evaluation in the organisation.”**
- **“Be more ‘sceptical’ during planning; testing and questioning proposals and for underlying assumptions.”**

## Development of the guidance

Some challenges around evaluation were discussed during the sessions which were taken into account during the development of the guidance.

### Measurability

Some participants seemed to feel strongly that qualitative data, or ‘soft outcomes’, should be considered sufficient or preferable to quantitative data in evaluation.

The guidance focuses on measurable, quantifiable, change, which inevitably means quantitative data. It may therefore be necessary to include a few short clauses in the guidance explaining the aims and limitations of this guidance and the reasons why qualitative processes are not covered.

### The evaluation needs of funders

For some participants, the reporting demands of funders were felt to prevent organisations from undertaking the kind of evaluation that they wanted. It is not inevitable that the reporting requirements of funders prevent appropriate evaluation. There may be situations where funder’s reporting requirements and the project’s ideal evaluation processes overlap to different degrees, but one does not necessarily prevent the other.

For the purposes of the guidance, this can be covered in a relatively simple statement that evaluation and reporting to funders are different processes (although there may be overlap), and that prioritising resources for evaluation is a matter for the organisation or project.

### Internal commissioning approach

There was substantial interest expressed in the idea of treating self-evaluation (evaluation processes undertaken by the organisation delivering the evaluation rather than an external party) in a manner similar to how an external evaluation would be commissioned. This was seen to have the potential to eliminate or reduce many of the barriers and challenges associated with self-evaluation.

For the purposes of the guidance, we may wish to give greater consideration to this as a technique promoted within the guidance.

## Conclusion

The complexity of evaluation practice, particularly in the field of anti-prejudice work, makes successfully delivering the required training a challenge. Despite this and the varied needs of participants, the feedback and experience of delivering the training suggests capacity building sessions like these may be suitable to support the guidance. However, this is unlikely to address the outstanding need for more in-depth, personalised support expressed by some participants during delivery.

These sessions also raised some surprising aspects; there was no observable relationship between the size and scale of the organisations represented and the prior understanding of evaluation of training participants. This contrasts with earlier assumptions that 'smaller organisations' may have a greater requirement for capacity building in evaluation.

There may be a case for investment in further training which allows a greater level of detailed discussion on topics which couldn't be included in these sessions. For many participants, interest was not for more detail on the topics covered, or for other topics, but for more specific support in enabling them to apply evaluation to their particular situation.

### Summary of barriers and solutions:

In each session, participants undertook an interactive discussion on barriers and solutions to evaluating anti-prejudice initiatives. Across the three sessions, the barriers raised were almost identical, and most of these fitted into one of the key overarching barriers already identified through the guidance development process.

Two additional overarching barriers arose through the group discussions:

- **Short-term project funding makes post-delivery phases of evaluation difficult or impossible to accomplish.**
- **Evaluation is considered too late in the project planning and delivery process to be designed effectively.**

The range of solutions raised also fell into a series of key areas:

- **Creating a culture within organisations and across funders where reflective evaluation focussed on 'what works' becomes the norm, and there is less fear of asking difficult questions / exposing weaknesses.**

- **Developing a consistent, proportionate approach to evaluation within the organisation, where strong evidence is used to develop the intervention itself and to build in robust evaluation from the outset.**
- **Specifying the approach to evaluation clearly (for example through an internal commissioning statement).**
- **Ensuring appropriate leadership on evaluation within the organisation.**
- **Developing partnership approaches between less well-resourced / less experienced organisations and those with more experience (for example universities).**
- **Sharing best practice and rolling out successful approaches.**
- **Ensuring that approaches to evaluation are appealing and accessible for those who are expected to participate.**
- **Embedding transparency and objectivity within the evaluation process.**
- **Exploring ways to map progress in a way that fits the organisation's desired outcome, rather than the expectations of funders.**
- **Providing training and skills development for staff / volunteers undertaking evaluation (preferably accredited).**
- **Adequately resourcing and investing in evaluation, including through realistic estimation of the cost, time and other resources required during project design and funding applications.**

The solutions identified here were built into the guidance in a more nuanced and flexible way, helping to shape the practical information and quality standards.

# Contacts

This publication and related equality and human rights resources are available from the Commission's website: **[www.equalityhumanrights.com](http://www.equalityhumanrights.com)**.

For advice, information or guidance on equality, discrimination or human rights issues, please contact the Equality Advisory and Support Service, a free and independent service.

Website **[www.equalityadvisoryservice.com](http://www.equalityadvisoryservice.com)**  
Telephone 0808 800 0082  
Textphone 0808 800 0084  
Hours 09:00 to 19:00 (Monday to Friday)  
10:00 to 14:00 (Saturday)  
Post FREEPOST EASS HELPLINE FPN6521

Questions and comments regarding this publication may be addressed to: **[correspondence@equalityhumanrights.com](mailto:correspondence@equalityhumanrights.com)**. The Commission welcomes your feedback.

## Alternative formats

This publication is also available as a Microsoft Word file from **[www.equalityhumanrights.com](http://www.equalityhumanrights.com)**. For information on accessing a Commission publication in an alternative format, please contact: **[correspondence@equalityhumanrights.com](mailto:correspondence@equalityhumanrights.com)**.

© 2017 Equality and Human Rights Commission

Published November 2017

Front cover image: [Shutterstock.com/Rawpixel.com](http://Shutterstock.com/Rawpixel.com)

You can download this publication from

[www.equalityhumanrights.com](http://www.equalityhumanrights.com)

© 2017 Equality and Human Rights Commission  
Published: November 2017

ISBN 978-1-84206-710-9

---