Using standards of evidence to evaluate impact of outreach

This guide has aims to help outreach providers to strengthen evaluation of the impact of outreach, in order to achieve good standards of evidence in impact evaluation. It aims to help providers to understand what constitutes high quality evidence and guide the selection of evaluation methods to generate evidence of impact. The document highlights practices that can strengthen the generation and use of evidence, and offers case studies and signposting to further sources.

The guidance is for people who already have some experience with evaluation techniques and are looking to make evaluations more robust and embedded.

The document has been developed with evaluation of the impact of outreach in mind, however many of the principles and practices will be relevant to other aspects of access and participation strategies.

This is part of a series of publications and should be read in conjunction with the following publications:

- Access and participation standards of evidence
- An evaluation self-assessment tool

OVERVIEW

1. Introduction	What is the guide and who should use it?
2. Setting the scene for evaluation: strategic level considerations	Support for evaluation Developing an evaluation culture Skills for evaluation
3. Designing your programmes	Rationale for programmes Using evidence to inform programme design Indicators and measures Building in evaluation
4. Evaluation design	What type of evaluation to aim for Discussion of different types of evaluation and research methods than can strengthen the evidence base
5. Implementing your evaluation	Developing your research strategy Collecting data Validating tools Tracking
6. Learning from evaluation	Use of evaluation Interpreting results Sharing results
7. Further support	Standards documents Evaluation resources
Glossary	

Style Features

Case Study:
Descriptive example of an approach or technique in practice.

Skill Requirements:
Planning, data analysis skills, communication

Further help:
Ideas about further reference sources.

Over to you:
What type of evaluation is it according to the standards?

Tips:
It helps if everyone working in the outreach team understands the importance of evaluation

1   Introduction

1.1   What is this guidance for?

This guide has been developed specifically to support the evaluation of the impact of outreach in UK higher education (HE), however the principles and practices are also relevant to other aspects of access and participation strategies. The Office for Students (OfS) wants to see strategic, evidence-led approaches, and good evaluation is key to achieving this. The Access and Participation (A&P) Plan asks you to demonstrate that you have a suitably robust evaluation strategy in place to measure the impact of your activities. Evaluation is important before, during and after planning outreach, and allows you to track and benchmark your results.

Good evaluation is needed for evidence-informed practice.

The aim of this guide is to show providers of HE outreach how they might be able to implement standards into their impact evaluation practice, with a focus on helping providers to select appropriate methods of evaluation, to improve the quality of, and usefulness, of the evidence generated, and to understand what claims can be made from different types of evidence.

The guidance shows how to apply standards to evaluation.

By sharing examples of useful approaches based on the experience of a range of different outreach providers, the guidance is designed to improve understanding of evaluation of the impact of outreach, enhance evaluation capability and promote better use of evauation to inform practice.

It gives examples of different practices.

The guidance is designed to assist providers in undertaking a self-assessment of their impact evaluation and develop their plans to improve their approach to undertaking and using impact evaluation. The guide reflects the dimensions of evaluation practice included in the OfS evaluation self-assessment tool.

The suggestions can support provider self-assessment and action planning.

1.2   Who should use it?

The guidance is for people who already have experience of evaluation techniques and are looking to make their evaluations more robust and embedded. It is suitable for people who want to know more about other practitioners' experiences of impact evaluation of outreach interventions, having already gained an understanding of evaluation techniques. Section 7 signposts to resources providing a general introduction for those who are new to the concept of impact evaluation.

This is not a 'how to' guide and therefore is most appropriate for people who already have an understanding of basic evaluation concepts.

This guidance is for use by all outreach professionals in UK HE. Third-sector organisations working in this field might also find it a useful reference point.

1.3   How should the guidance be used?

You can use the document to promote a culture of evaluation across a range of activities. You can also use it to develop an evaluation strategy with one particular activity in mind, and as a framework for evaluation of the overall outreach programme strategy.

The guidance is designed to give ideas and practice examples to help your thinking on how your own practice could be strengthened. The guidance is organised according to the dimensions of the evaluation self-assessment tool, and can support your self-assessment of what you are already doing to generate high quality evidence of impact and provide suggestions and examples to inform your future evaluation action plans and approaches.

The concepts in the guide can be applied to specific projects or whole programmes of activity.

The guidance includes reflective questions and the skill requirements, which are designed to be of use to help to inform your impact evaluation plans. This guidance does

not provide 'how to' tools for evaluation, although it does signpost to further resources on implementing different types of evaluation approaches.

1.4   Why are evaluation standards important?

Evaluation is an integral part of any programme of widening participation outreach because it is vital to be able to demonstrate the impact of outreach activities and to ensure the work is effective. Evaluation is also important to inform the on-going development of outreach: i.e. not just to 'prove' but to 'improve'. Significant expenditure is allocated to outreach each year and decisions on resourcing should be based on sound evidence of effectiveness. Higher education providers and others delivering outreach have a responsibility to progress evidence-based interventions: either based on evidence from your own evaluations or from evaluations of effective practice elsewhere.

Evaluation supports effective practice by enabling providers and the sector to identify what works well and what needs improving.

Tip: Ensure that everyone working in the outreach team understands the importance of having evidence-informed practices.

2   Setting the scene for evaluation: strategic level considerations

The strategic context for widening participation outreach varies across different providers, for example, responsibility could sit within a widening participation department, or with admissions or marketing teams. Wherever responsibility is located, there are several ingredients to ensuring that impact evaluation is taking place as it should be and moving towards an increasingly sound and robust evidence base, as discussed here.

2.1   Support for evaluation

Evaluation should be prioritised alongside planning, project implementation and documentation. This means developing regular opportunities for access and participation staff to talk about evaluation as part of the day-to-day work. Ideally there should also be strategic overview of what evaluation is being undertaken, so that the implications of the findings for institutional practice can be discussed.

Institutional opportunities and structures are needed to discuss evaluation of impact.

Senior leaders are well placed to plan and allocate the significant resources required for evaluation and to support the development of an environment in which evaluation is talked about regularly. It is crucial that senior leaders and heads of widening participation buy into the importance of undertaking robust evaluations. Ultimately it would be desirable for outreach providers to view evaluation as a strategy – i.e. part of the institutional approach to outreach and built into the fabric of what is delivered. Senior buy-in is required in order that institutions do not shy away from evaluations that might lead to difficult conversations about the quality and effectiveness of the work being done.

Senior managers should understand the importance of embedding evaluation.

You should aim to plan for evaluation activities that are of strategic relevance for the overall outreach programme. Having evaluations that complement each other across the range of activities allows you to compare and contrast the results. It makes sense to use similar indicators of success which link to overall strategic aims and facilitate comparisons (i.e. a 'joined-up' approach). Working towards a 'whole-institution' approach of evaluation can often be valuable to take account of the high turnover of staff in the sector.

Evaluation that supports the strategic objectives should be prioritised.

It is also really crucial that senior leaders buy into the importance of doing evaluation well as this is going to require a commitment to ensuring that resources for evaluation are in place (time, money, expertise for evaluation).

Commitment to resourcing evaluation is required.

Over to you: Support for evaluation
Are there opportunities for your widening participation team to have conversations about evaluation on a regular basis? Is there senior level buy-in and a mechanism for strategic overview of evaluation of outreach? Are institutional resources deployed with evaluation aspects in mind?

Suggestions for practice:

Opportunities within teams to talk about impact evaluation can be created by, for example, making evaluation a regular agenda item at team meetings, or establishing forums with a remit to discuss evaluation activities and keep team members informed.

Maintaining a strategic overview of evaluations usually requires formal bodies to be set up, for example, within the institutional committee structure. In order to inform evaluation design and delivery it is often useful to involve members with a range of skills and perspectives such as practitioners, academics and students.

A 'joined-up' approach to evaluation can be supported by using common evaluation protocols or frameworks for evaluation of the impact of interventions. For example, some providers have put in place a common framework of outcome measures (where appropriate) across different programmes. Using shared templates for planning evaluations might also be another way to gain more consistency in how evaluations are approached and understood across the gamut of access and participation provision.

Commitment to securing appropriate resources for impact evaluation can be demonstrated by, for example, making sure that project and programme budgets include a budget line for evaluation. You might consider setting a baseline for the share of overall costs that are allocated to evaluation (i.e. as a proportion of the access and participation activity).

Practice example: Developing internal structures to support A&P planning and evaluation

For new providers that are starting out on developing their access and participation strategies, impact evaluation should be seen as integral to taking forward new initiatives. Ideally roles and responsibilities in terms of agreeing, implementing, managing and learning from evaluation will be in place from the start, alongside formal structures to ensure that the delivery of outreach develops in a way that responds to evidence about what works best in their particular context to maximise the outcomes and impact of the access and participation investment.

Case study: London School of Management Education (LSME)[1]

Background: The college is developing a strategy for access and participation in a three-year cycle, drawing on a college-wide approach that joins the expertise of academics and professionals in order to evaluate the feasibility of the proposed strategy and capture the potential outcomes and impact.

How tackled: A series of informal conversations about the access and participation strategy with academic and non-academic staff included discussion of an appropriate evaluation of its success by the end of the planned cycle. An access and participation working group was established. Headed by the executive director, the group comprises the chief research and innovations officer, two academic staff with a research background, the partnership development officer and a student representative. This working group has the overall responsibility of evaluating the outcomes and changing the course of action if required. The members of the working group have varying expertise, although all are aware of the importance of evaluation and are committed to facilitating robust data collection processes. The group has developed a common protocol for evaluating access and participation projects with common measures of success and outcomes across the entire teaching and learning activities in the whole institution. The OfS draft self-assessment tool for evaluation provided a broad framework for reflecting on LSME's future evaluation of the access and participation activities.

Results and learning: The activities and associated evaluation will formally commence in March 2019. However, even before the planned cycle of the current strategy there have been positive benefits because having a multi-disciplinary team presented a valuable opportunity to tap into each other's expertise. For

---

[1] LSME is a small alternative provider of higher education, relatively recently involved in access and participation planning processes, as a mandatory requirement of OfS registration.

example, it was clear that LSME's proposed collaborative partners who have been delivering outreach have similar affiliations with several organisations, meaning that the developing relationships will be essential to success. To strengthen these relationships, LSME formally engaged potential collaborators in internal and external activities and contributed to the development of their other initiatives including participating in job and career fairs to promote each organisation and facilitate sharing of knowledge.

Despite the possibility of constraints in understanding the roles of each individual, the small size of the institution promotes informal dialogue which helps to build up professional relationships among colleagues. The self-assessment tool has helped to identify areas for development to enhance the evaluation of the outcomes of the access and participation outreach activities to align with the Standards of Evidence. For example, participant data is collected at the beginning and during the period of evaluation making it practical to consider individual change that could be a direct result of the interventions, and the evidence-base draws on data from different sources (participants, staff, partners and other stakeholders).

How could this approach be developed further? A key challenge for the group will be their ability to conduct robust research with limited funding and simultaneously produce reliable baseline findings for future evaluations. The process of identifying the right skills for evaluation is being discussed by the working group. The identified skills will be matched to the existing skillset of staff and responsibilities will be assigned to ensure an effective evaluation. The information on the evaluation approach and the specific roles of key staff will be actively disseminated across the institution and to collaborating partners to facilitate their maximum engagement with the evaluation process.

2.2 Developing an evaluation culture

An organisation with a culture of data and learning has greater capacity to benefit from evaluations. It becomes a learning organisation, in which leadership and staff continually improve upon ongoing programmes and develop their abilities to achieve the results desired.

Impact evaluation will be most effective in organisations that are open to learning and create opportunities for outreach practitioners to reflect and develop.

A learning organisation can be built by encouraging practitioners to develop reflective practices, by modelling good evaluation, by asking questions about the linkage and availability of data and information systems, and by using data to make informed decisions. It is helpful if some staffing resource is dedicated to evaluation or at least to have staff time protected to undertake evaluation work.

Tip: Work towards everyone understanding the importance of evaluation. Not only the outreach team but the institution as a whole, including professional services staff, academics and students

Over to you: Building an evaluation culture
Are outreach delivery staff and partners aware of the importance of evaluation and committed to facilitating robust data collection processes? Do you create opportunities for honest reflection on the effectiveness (or otherwise) of your activities? Is there a whole-institution approach?

Practice example: Regular review cycles

Setting up structures to promote regular and ongoing discussion of impact evaluation, such as working groups or committees, is one way of promoting a culture of evaluation, and might be especially important within larger institutions. These can also help when it comes to learning from the results of evaluations and considering the implications of the access and participation strategy. Thinking about cycles of evaluations that fit with the cycles of programme planning is a worthy goal that will help to ensure that evaluations can inform the evolution of programmes and how outreach is conceptualised and delivered.

Case study: University of Liverpool 'Scholars' Programme

Background: Previous evaluation suggested that the Liverpool Scholars outreach programme helps students to have realistic expectations of the demands of being a student at the University of Liverpool; establishes a sense of relationship with the university ahead of taking up a place; and supports the development of

academic skills. From the testimony of Scholars students this journey was characterised by a process of personal change as the young person made a transition from identifying with the university (as something to which they might aspire) to identification as a student at the university (as something that they would achieve). The Scholars team wanted to build upon this work and to deepen their approach to evaluation by adopting an 'embedded approach' involving planning cycles of evaluation activity that build year-on-year through consecutive rounds of professional reflection, evaluation activity and programme review.

How tackled: An evaluation lead was established for the programme. A workshop event was organised, attended by the Scholars team, and other staff with relevant roles. At this event, changes in the conceptualisation and design of the programme were reflected upon. The widening of the design beyond academic skills to include preparation for student life, coping skills and resilience was discussed. The outcomes from this professional reflection fed into the design of a questionnaire survey sent to Year 1 undergraduates who had come through the Scholars route. The focus of this survey was the experience of the students post-entry (whereas previously the focus had been upon the approach to taking up a university place). In addition, once outcomes data for Scholars students was available, including Year 1 to Year 2 progression outcomes, it was analysed to compare against the 'all undergraduate average' and the average for students in receipt of bursaries.

Results and learning: The survey data suggests that students coming through the Scholars route were well prepared for becoming undergraduate students. The responses revealed very high, positive ratings for questions concerning being able to cope with the pressures of academic life; having the skills needed to succeed on course; having realistic expectations of the academic demands; realistic expectations of student life; being able to take part in co-curricular activities within the academic department; and in extra-curricular sporting and/or guild/society activities. The analysis of on-course progression outcomes showed no significant difference between Scholars students and all other students. Therefore, the team concluded that participation on the Scholars programme had produced a levelling effect with respect to students from non-disadvantaged backgrounds, given students who had come through the Scholars route had experienced social and/or education disadvantages. The data on progression outcomes strengthened confidence in the evidence for the benefits of the Scholars programme that had been produced by the student survey.

How could this be developed further? The Scholars team at the University of Liverpool are working towards an embedded model of evaluation and reflection that is built into regular professional cycles throughout the academic year. Each cycle involves a critical professional reflection on key design features; collation of student perspectives on their experience and the benefits; and analysis of outcomes data (including Year 1 to Year 2 progression data). This process, iterated annually, aims to ensure a strong impact evaluation which is focused and purposeful.

2.3   Skills for evaluation

Evaluations usually require significant input to ensure they are designed and delivered successfully. High level expertise is most needed during the evaluation design stage, data analyses and reporting stages, but there will be an ongoing requirement in terms of data collection and the skills required to do this well should not be underestimated. Partnerships between academic staff and practitioners can be a way of drawing in appropriate expertise, if this is not available or can't be developed with the access and participation team.

A range of skills at different levels are required.

You have a choice of commissioning an external evaluation or undertaking evaluation work in-house. Both options have resource implications. The decision is likely to depend on the existing level of knowledge and skill and time available you have yourself or within your team.   It may also depend on the extent to which you are seeking external verification of the claims you are making. Getting an independent individual or organisation to undertake the evaluation is usually considered to lead to more objective results. Even if your evaluation is externally commissioned you will generally still require an appropriate

Commissioned evaluations can be used to bring in expertise and gain an external perspective.

internal project manager with the relevant skills to oversee the evaluation, and it is usually helpful to draw on a steering group of people with relevant insight and experience to govern the evaluation.

The decision about putting in place external or internal evaluation might not be an either/or choice. You could undertake some aspects within your organisation and outsource others or ask for expert advice or support for more complex aspects of the work. For example, you might have focus group data transcribed or you might ask an expert statistician regarding your data analysis. You might also outsource some evaluation projects entirely and conduct others yourself (however even if you commission research, you should understand how such research maps onto the types of evaluation).

There are no hard and fast rules about who should undertake evaluation, although those involved should be objective.

Skills required:

Different skills will be needed depending on role:

Project management: development of the evaluation specification; development of research instruments; implementation of data collection, storage and data protection; external commissioning (if required); data analyses and interpretation; development of findings and conclusions; preparation of final reports and presentations.

Senior management: prioritisation and resource allocation; management of risks; quality assurance of the evaluation design, research instruments and outputs.

Steering group: developing the evaluation questions; design of evaluation methods; access to information and contacts.

Over to you: Skills for evaluation
Have you identified a skills base/expertise among professional service staff for undertaking or commissioning evaluation of widening participation initiatives? Have you identified a skills base/expertise among academic staff for undertaking or commissioning evaluation of widening participation initiatives? Are there opportunities for staff to enhance their evaluation skills and understanding?

Suggestions for practice:

In larger institutions with a research culture it may be possible to identify staff with expertise who can support evaluations, for example through expertise in evaluation techniques or data analysis. Small institutions without a tradition of research may need to utilise external consultancy.

All providers might consider joining local consortia for undertaking evaluation, or networks for sharing of information on effective access and participation evaluation practices.

If there is sufficient resource available, institutions could consider employing an evaluation team member with a remit for supporting the design and delivery of evaluations of impact and effectiveness.

Many evaluation training materials and resources are available, as well as events and networks. Prioritising relevant training/up-skilling opportunities is a useful way of improving the team skillset in relation to evaluation. For example, this could be part of team members' regular personal development review process.

Practice Example: Practitioners and academics joining forces

Opportunities for joining up the expertise located among widening participation professionals and academics was identified as a means to developing evidence-based practices for widening participation students. Practitioners and academics want to draw on each other's expertise to ensure that access and participation activities are impactful and evidence-led.

Case Study: University of Exeter

How the initiative developed: The head of widening participation and an associate professor in HE jointly founded the Centre for Social Mobility (launched in June 2018). The centre aims to combine practice and research into supporting social mobility through HE, for the benefit of students, prospective students, staff and partners, and specifically, to develop and undertake institutional and sector research that enhances practice. The approach is designed to embed a whole-institution approach to widening participation.

The first stage was to map the abilities and roles of different members of staff and to develop reach into relevant committees with power to shape and influence the social mobility agenda. This has laid the foundations for conversations and scoping work to develop evaluations, and enhance practices including tightening up the data systems. The process has been partly 'top-down' – driven by the strategic requirements and A&P Plan process – and partly 'grassroots' – highlighting issues and challenges on the ground. There are plans to evaluate the institution's employability services, and the widening participation team have helped to shape the development of a new research project in the School of Education.

The initiative has also involved drawing in additional expertise and human resources for evaluation. A Masters student is analysing data collected from the Realising Opportunities programme on aspirations for medicine in a way that practitioners may not have the time and resources to do. The findings of this evaluation are designed to feed back into practice. The centre is submitting bids for external funding for research and evaluation that combine the expertise of professional service and academic staff, a process which those involved find mutually enriching and beneficial.

To facilitate a culture of dialogue the centre held an internal social mobility conference, which showcased expertise and projects from professional services, students and academics. A seminar series is also underway with speakers who are relevant to academic researchers and professional services staff members working in social mobility.

Results and learning: It is a journey for colleagues with very different role requirements to work collaboratively and to understand the opportunities and constraints within each other's roles and it takes time and good will to develop genuine collaborations and to start joint endeavours. However, fostering dialogue is key, including physically connecting colleagues. A key challenge for academics is how to seek to conduct rigorous research within an institution whilst simultaneously meeting the requirements of the external research environment (such as the Research Excellence Framework (REF)).

How could the approach be developed? Further work could be done to map existing evaluation practices against the types of evaluation, and to create and sustain a culture of evaluation and reflection. It will be desirable for academic and professional service staff members to take forward joint initiatives to review existing evaluations and to develop and implement new evaluation approaches to widening participation initiatives spanning the entire student lifecycle.

3 Designing your programmes

Programme design consists of identifying solutions to an unsatisfactory situation, and putting in place plans to deliver activities to achieve desired results. Considerations at the programme design stage help to lay the foundation of good evaluation and accountability, as discussed here.

3.1 Programme rationale

The first step in programme design is to gain clarity over the issue that your programme is addressing (i.e. being clear about the nature of the problem you want to resolve and why). You should think about who the problem affects and why making an intervention to address this matters for HE progression outcomes. As well as helping to define the programme rationale and your objectives for the intervention, clarity about the underpinning issues you want to address is key to building a shared understanding of your

Programmes should be underpinned by clarity on what

programmes. Without taking time to actively describe and document these things it would be easy for the specific issue to remain implicit in the assumptions.

you want to achieve.

Part of this is also about 'sense-checking' the intervention, i.e. clarifying the issue is pertinent to your objectives and that your planned activities are capable of making the changes to want to see. Techniques like Theory of Change mapping which identifies the processes which are anticipated to lead to the desired outcomes (see below) will help you to think through the factors that might support or undermine what you are planning.

Be clear on the activities and the rationale for delivering them in this way.

Why you are delivering activities in a particular way is the most important question when developing an intervention and is at the heart of outreach and evaluation. It moves you from the practical concerns about organising and implementing an intervention to more significant questions about what you think are the underlying principles which are causing the change you want. Whether you are starting out on a new outreach project, or you have been running an intervention for a long time, you should prepare the ground for your evaluation by setting out a coherent description of what you are doing and why, and your measures of success, as well as working out the most suitable methodology to use to show the impact of what you do.

Develop a coherent description of what you are doing and why.

Tip: The 'issue' could be a specific problem (e.g. lack of information or specific skills) or a general negative situation (e.g. low levels of motivation or coping). You should specify the improvement that can be achieved by the outreach intervention and why this is important.

There are various tools and also a plethora of terms that have developed in association with programme theories (for example: Theory of Change; logic models; outcome chains; logframes) but do not be worried by these. If you are using a Theory of Change for the first time, keep it simple and clear, and then increase complexity as you become more familiar with the approach. As you do this, remember that it is the thinking and the sharing of ideas about how your intervention will create change that is most valuable: any diagram or narrative account is a representation of that thinking.

Various tools can be used to set out the theory underpinning your approach (the intervention theory).

In summary, tools such as Theory of Change and logic models strengthen evaluation because you are clearer at the outset about your goals, the pathways to achieve those goals, and the causal relationships that are driving change. From this you can develop a robust evaluation framework, which considers what will be the best data/measures to evaluate the intervention effectively.

Focus on causal relationships to achieve your goals and how to measure results.

If you look at the resources below or explore the web you will find a huge variety of different ways of presenting your intervention theory, some of which are highly complex. To start with, keep things simple for your first attempt. You can then make them more complex in the future if it helps. Some people prefer written narratives to diagrams because they allow more space for articulating the causal relationships. But always remember, the real benefit is not the complexity of the resulting model, but the quality of the thinking the process demands.

The quality of thinking is more important than how the theory is presented.

Skills required: Understanding of the policy context in which you operate nationally, locally and within your institution; literature reviewing and synthesis; understanding social theory; understanding previous research; group facilitation; sharing and promoting a joint understanding of your intervention theory.

Tip: Focus on the intended outcome first and design your intervention to achieve that outcome, not the other way around. The important thing is the quality of thinking; avoid getting too drawn into complex diagrams.

Over to you: Why do you do it?
Can you demonstrate understanding of what you are trying to achieve and how you will measure success against your goals?
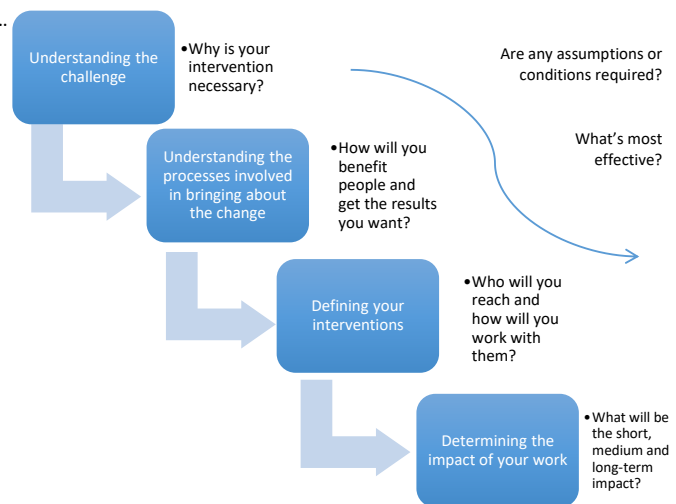
Practice example: Developing a Theory of Change

One increasingly common way to explore the assumptions and processes underpinning your delivery is to develop **a Theory of Change**. Theories of Change have been successfully used by not-for-profit and charitable organisations to help them evaluate their work but are now becoming much more common in educational settings.

Theory of Change is a way of thinking strategically about a desired change and how you will achieve it. Typically you will create a diagrammatic representation (sometimes called a **logic model**) of your theory which can act as a reference point throughout your intervention. This will show the pathway of change from the inputs to the outcomes you hope to achieve.

Often outreach interventions begin with an idea for an intervention and everything is planned from that point: in other words the activity is the dominant focus not the processes of change. Theories of Change reverse this and invite you to think about the change you want to effect, and the nature of the problem and only then about what actions might achieve this. For example if you have identified low expectations to progress in education as a barrier to university access the next step is to think about what processes might change this. Don't be surprised if this is quite challenging – you are trying to dig deep into the problem and analyse change processes which may not be a quick thing to do. Usually working to address issues of access and participation in HE involves working in partnership with others, so reflect on who else you need to involve in developing your intervention. To develop and embed your Theory of Change it is advisable to work with others in your organisation and external stakeholders including students and those who advise them.

The main stages in developing a Theory of Change are:

Start here…

- Understanding the challenge
  - •Why is your intervention necessary?
- Understanding the processes involved in bringing about the change
  - •How will you benefit people and get the results you want?
- Defining your interventions
  - •Who will you reach and how will you work with them?
- Determining the impact of your work
  - •What will be the short, medium and long-term impact?

Are any assumptions or conditions required?

What's most effective?

Developing a Theory of Change is first and foremost **a process of thinking** through an intervention in depth from conception to its conclusion. It is a way of shaping your thinking and planning a kind of mapping exercise both mentally and on paper. The discussion begins with the desired goal – what do you want to achieve – and then plots what will help you get there and what might act as barriers and constraints. An advantage of developing a Theory of Change is that it can bring a widening participation team together to develop shared understanding of the intervention which brings more critical engagement with its implementation and evaluation.

Another important aspect of a Theory of Change is that helps you to explore **causal relationships** – what it is in your intervention that might be causing the changes that you want to secure. With widening participation interventions it is all too easy to become more focused on *what* you will do rather than on *why* it might work, and to be more preoccupied on the activity than what it is which is actually driving the change. Understanding causal relationships is critical for two reasons. Firstly, if you do not know what is causing the change you are trying to achieve, then it is very hard to refine and improve your intervention to make it more effective. Secondly, if you do not know what is causing the change, it is hard to generalise or upscale from your intervention.

Case Study: Coachbright[2]

How tackled:

Coachbright's Theory of Change was developed with help from an external agency familiar with Theory of Change models. In their diagram you can see the intervention described in the Inputs section, and careful thinking about the short, medium and long-term outcomes. The outcomes are sub-divided into themes: Academic Achievement; Confidence and Self-Esteem; and Aspirations and Expectations. The detail in outcomes will help in ascertaining whether some outcomes seem to be more or less affected.

Result and learning:

This is an example of constructive practice: to work with someone more experienced initially and to co-create the Theory of Change diagram. This has lead to better understanding and will enable Coachbright to articulate their programme theories in the future.

How could this be developed further:

Some key evaluation questions that Coachbright could ask, prompted by the Theory of Change, might be:
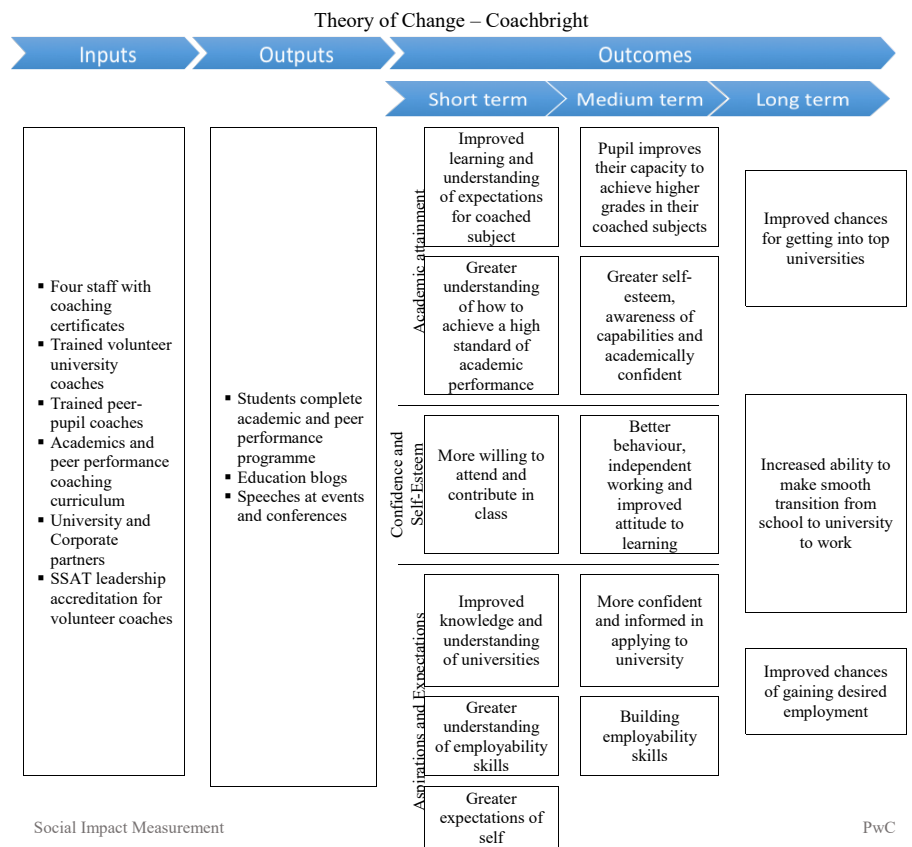
Are we clear enough about the precise nature of the learning we hope will be improved?

What are the employability skills we hope the students will understand better?

What is the difference between confidence and self-esteem?

How can we measure confidence and 'informedness' in applying to university?

How can we capture the long-term outcomes data?

Theory of Change – Coachbright

Inputs	Outputs	Outcomes		
		Short term	**Medium term**	**Long term**

Academic attainment

| | | Improved learning and understanding of expectations for coached subject | Pupil improves their capacity to achieve higher grades in their coached subjects | |
| | | Greater understanding of how to achieve a high standard of academic performance | Greater self-esteem, awareness of capabilities and academically confident | Improved chances for getting into top universities |

- Four staff with coaching certificates
- Trained volunteer university coaches
- Trained peer-pupil coaches
- Academics and peer performance coaching curriculum
- University and Corporate partners
- SSAT leadership accreditation for volunteer coaches

- Students complete academic and peer performance programme
- Education blogs
- Speeches at events and conferences

Confidence and Self-Esteem

| More willing to attend and contribute in class | Better behaviour, independent working and improved attitude to learning | Increased ability to make smooth transition from school to university to work |

Aspirations and Expectations

Improved knowledge and understanding of universities	More confident and informed in applying to university	
Greater understanding of employability skills	Building employability skills	Improved chances of gaining desired employment
Greater expectations of self		

[2] Coachbright is a social enterprise that exists to support pupils from low-income backgrounds to become more independent and resilient so they can lead the lives they want. The organisation runs academic coaching programmes that support pupils to improve their grades, confidence and expectations.

3.2 Using evidence to inform programme design

Evidence about existing providers' practice in relation to the issue you are addressing and the impact of different approaches on the target group's journey to HE (if any evidence currently exists) should be used to inform your programme design decisions. Different scenarios can be envisaged:

Your programme design should be informed by existing evidence of what works. If no evidence is available then you'll need to build in evaluation in order to generate evidence to inform how the activity is developed.

- The body of convincing evidence suggests the programme and activities you are planning have beneficial results, linked to a specification of approach that is transferable to your particular context, i.e. evidence supports the 'business case' for using resources to take forward the planned interventions and provides pointers to practices that work.

- There is little or no robust evidence to show the benefits of the intervention (this could be in relation to participants as a whole or the specific sub-groups you are interested in). In this scenario the evidence establishes that it is appropriate to conduct an evaluation study to test the effectiveness and impact of your planned activities. This might also be the case if the evidence suggests some uncertainty about the comparative benefits of alternative interventions.

Evidence includes the results of your own and other people's impact evaluations. For example, taking account of published research and national data, along with monitoring evidence, and your own previous evaluation(s). It should be noted that the approach described here is non-systematic in that issues associated with the reliability of the evidence and publication bias are not addressed. Undertaking a systematic approach to reviewing evidence comprises a Type 1 evaluation in its own right (as discussed below).

Other people's evidence of what works is useful if robust and transferable to your situation.

The important thing is to ensure that you understand what the existing evidence says about the likely results of different approaches and the implications for your programme design. This could be an iterative process: as you consider the existing evidence you should refine your ideas about the intervention you want to test, and as the programme design is developed you may need to search for more evidence relevant to the specific design features. In some cases it might involve a process of choosing between different courses of action, especially where time and resources (material, financial, human) are limited.

Use of evidence should inform decisions on the course of action to take.

Tip: Using evidence on an intervention is intended to support making a judgement along the lines of: 'is this intervention likely to be effective in addressing the particular issue and context I am interested in?'

Over to you: Why do you do it?
Is your programme informed by the evidence?

Potential approaches to gathering Type 1 evaluation evidence include:

Type	Sources
Evaluation results	Drawing on the results of your previous evaluations of the programme in question, or other previous similar activities.
Citations and references	Review of relevant theoretical or practitioner literature, including scholarly literature as well as government and other reports.
Knowledge sharing	Participation in conferences or other types of engagement and knowledge sharing with other outreach practitioners at regional, national or international level.
Literature reviews	Evidence of keeping continuously up-to-date, including review cycles for renewing literature reviews.
Call for evidence	Putting out a call for evidence by getting in contact with practitioners and other experts who have been involved in evaluation of outreach interventions.
Repositories of evidence	Submitting research findings and evidence from evaluation to the Evidence and Impact Exchange (EiX) so it can be synthesised, translated and disseminated to support knowledge exchange to a wider audience. Further details at: https://www.officeforstudents.org.uk/advice-and-guidance/promoting-equal-opportunities/using-evidence-and-evaluation-to-improve-outcomes/evidence-and-impact-exchange/

3.3 Indicators and measures

Complex challenges like the underrepresentation of disadvantaged students in HE have multiple causes and many different implications depending on the context. Sometimes the initial problem may only be a symptom of a deeper issue and solutions may not always be clear. Although the 'big picture' goal of widening participation is an underpinning aim, most interventions are addressing a smaller element of that big picture goal. The ultimate outcome of increasing the number of disadvantaged students benefitting from a university education is likely to be the consequence of smaller, more focused interventions that identify and address one aspect of the journey to university. For example, interventions may be setting out to: realise aspirations; boost examination results; develop familiarity with life at university; build self-efficacy and resilience and so on. The indicators and measures you will use to evaluate the success of programmes need to be identified at the programme-design stage in order to ensure processes are built in to collect the data required.

Be clear on how you will measure all of the outcomes and impacts of your activities at programme design stage.

Key Terms

Outcome: Measure of the positive changes your activities are making to those who take part (pre and post)

Impact: Measure of the difference you are making to HE access and participation

Once you have articulated your intervention theory, with accompanying clarity about intended outcome indicators, it becomes much more straightforward to develop an evaluation plan and a purposeful evaluation of your intervention. The intended outcomes are the starting point for thinking about impact evaluation as you are trying to determine

The outcomes you want to achieve are the starting point for thinking about how to evaluate.

how effective your intervention has been in achieving these. You also need to consider the relationship of the outcomes to the longer-term impacts of your work.

Once the aims of your outreach project strategy have been comprehensively defined, they need to be matched to indicators against which outcomes and impacts can be assessed. Indicators are measures of the change you are attempting to bring about – i.e. capable of capturing the impact on individuals, groups, organisations or systems. The indicators should be specified at the design stage to ensure that methods and processes are put in place for the collection of the data relevant to inform the indicators. Indicators are different to objectives in that they have certain important characteristics, namely they are specific, measurable (i.e. through applying appropriate data collection methods), and should directly relate to the achievements of the intervention you are making. If no aims or no outcome indicators have been identified or if data cannot be collected, you should question why you are conducting this outreach activity.

You will need to identify 'indicators' capable of capturing the changes you are looking to achieve.

Tip: Create a list of outcome indicators prompted by your Theory of Change. Make sure you use indicators that directly reflect your interventions and the aims of what you are doing.

The following factors should inform your choice of outcome indicators:

Outcome indicators should be relevant to what you are delivering and why you are doing it.

- Relevance to the activities (i.e. outcomes that directly relate to the approach and practices you are adopting should be favoured);

- Relevance to the objectives (i.e. outcomes that represent the underlying aims of what your programmes seek to achieve);

- Availability of data (in some cases direct measures of an outcome might not be available and you will need to use a proxy or surrogate, however, direct measures should be favoured where possible).

Direct measures are better than proxies.

You should consider the relevance of the outcomes you are measuring to the access and participation objectives. That is, you should be able to demonstrate that where a positive change occurs that this signals an improvement in HE participation prospects. For example, levels of attainment in exams could be a seen as a predictor of HE participation (if your project is designed to improve attainment). The main thing is to choose success measures in terms of achieving outcomes for participants (i.e. moving beyond feedback/satisfaction measures or the opinions of participants e.g. continuation and progression, attainment, behavioural changes). If the focus of your evaluation is on 'raising aspirations' for HE consider expanding your questions to take in the expectations of the young person as research suggests these have stronger predictive power. If you are already asking young people about their expectations around HE, consider broadening out the questions to take in what they think their parents and teachers expect as research suggests that these all have a strong correlation with future behaviour.

Consider the relevance of the data to future behaviour and HE participation.

The existing evidence on interventions might also be an important source when considering the potential outcome measures for your programmes (i.e. it's useful to make reference to the existing evidence base to identify outcome indicators that used in previous studies). This can help you in interpreting the results from your evaluations by benchmarking your outcomes against what other interventions have achieved.

Using common indicators can help you to compare results.

There are various ways of quantifying the results from outreach. Practical considerations usually come into play in the context of outreach, for example, the amount of time to collect a large amount of data from participants might be quite restricted apart from during intensive or repeat activities. Surveys are a way of collecting data at scale and can provide rich insights; however, use of linked tracking data might also be needed to make sure your data collection covers the outcome indicators relevant to HE access.

Practical considerations often influence what data is collected.

At least one outcome measure should be selected (and data collected on this). A hierarchy of outcome measures could be identified (i.e. to measure the primary outcomes and secondary outcomes, and any intermediary outcomes). You might also consider testing if there are any adverse outcomes (i.e. measuring any potential unintended or adverse effects).

Aim to collect data on at least one core outcome.

Tip: If no objectives have been set, or no outcome indicators identified, or if relevant data is not being collected, you should question why you are conducting this outreach activity.

The timescale of the evaluation needs to be carefully considered: the impact(s) may take years to materialise. In this case it is important to build in collection of data on the outcomes that can be used to measure the benefits in a shorter timeframe. Considering shorter, simpler 'links' in the logic chain can increase the ability of evaluations to provide good evaluation of impact if successfully converted into robust data collection measures. At the same time taking a long-term perspective and making a commitment to evaluating the outcomes of your evaluation longitudinally is going to be key to establishing whether it works.

Look at intermediate as well as final outcomes especially where these will take a long time to materialise.

It is often helpful to articulate **short, medium and long-term outcomes**. Typically, it is the long-term outcomes that link directly to the aim of increasing access and participation in HE, whilst short/medium-term outcomes are more specifically related to the objectives of your outreach. A benefit of this way of thinking is that it recognises that resolving the widening participation problem is unlikely to be a simple quick fix but a combination of interventions, probably at different time points in a student's educational journey. It also encourages us to see how our interventions form a pathway to the goal of increasing participation, reminding us that a successful short-term outcome of an intervention may or may not translate to long-term outcomes.

You could consider a framework of indicators of outcomes at different stages in the learner journey.

Tip: By shortening the timescales for measured outcomes, the use of intermediate steps helps you to make stronger causal claims about outreach activities than those provided by very long-term perspectives where multiple confounding factors make it harder to disentangle influences on young people's decision making.

There are different ways of thinking about the outcomes from outreach over time, although a fairly common approach, which has gained currency amongst widening participation practitioners, is the Kirkpatrick approach. This model is designed to capture four dimensions:

The Kirkpatrick model helps you to identify outcomes at different levels.

Level	What is measured? (how)
Reaction	How participants feel about the experience (e.g. through feedback surveys and observations)
Learning	Increase in people's knowledge and skills (e.g. through formal and informal assessments of understanding, knowledge and skills before and after the intervention)
Behaviour	How far learning is applied and results in personal change, for example a decision to apply to HE/take up a place (e.g. through follow-up, follow-up over time, or tracking progression and attainment outcomes over time)
Results	How far the outreach impacts at a structural/societal level or organisational level, e.g. to identify whether participation has widened as a result of the intervention (e.g. analysis of outcomes of cohort using administrative datasets)

A further benefit of thinking about outcomes in this way is that it may prompt you to reflect on the **assumptions** that underpin your intervention theory. For example, the assumption that a one-off highly positive experience (such as a master class) which has been seen to

Breaking down the outcomes into stages can help you to reflect on the

successfully raise motivation for learning in the immediate term will lead to increased attainment at school in the short term and a raised sense of possible achievement may be over-optimistic. This, in turn, may prompt you to re-evaluate your intervention in terms of how sustained an experience it is. Or it may prompt you to look at how the school could help by, for example, re-enforcing the learning through other school-based activities.

assumptions and conditions by which an intervention is effective.

It is common for outreach evaluations to use self-reported data (for example, questionnaires with participants regarding their attitudes towards progression in education). However, self-reported data tends to be rather unreliable and unless there is very good questionnaire design there are likely to be validity issues. There are several potential problems: participants' responses may be inaccurate (for example, if they did not engage fully with the questions or misunderstood); responses may reflect participants' expectation about what they are expected to say rather than true beliefs; responses may be over-optimistic because an immediate reaction captured in a questionnaire at the end of the activity may not be sustained over time, or because respondents over-estimate their level of knowledge (the Dunning-Kruger effect). The results of self-reported surveys may be biased depending on who takes part since certain demographic groups have been shown to have a tendency to respond differently than others (for example, boys tend to rate their confidence higher than girls). All these issues mean that any claims made from self-reported data need to be carefully considered, and based on the best possible survey design. Self-reported data by its nature is subjective; it does not measure the occurrence of any concept as such (e.g. self-confidence) but rather what respondents' say based on a subjective assessment of themselves. Ideally self-reported data would be triangulated against other sources (e.g. teacher feedback or use of objective tests).

Drawing conclusions from purely self-reported data is problematic due to subjectivity issues.

Over to you: Capturing the activities and selecting indicators
Have you defined and agreed the deliverables for your programmes?   Are you clear on who will take part and what will be delivered? Are you clear on how you will measure your outcomes and impacts? Are your success measures focused on objective measures of the outcomes for participants (i.e. moving beyond feedback/satisfaction measures)? Can you point to evidence underpinning your choice of outcome measures for your access and participation programmes? Have you made sure you can identify how you will access the data required to measure outcomes and impact? Do you have benchmarks to measure your performance against?

Practice example: Developing measures of success

Well-developed thinking recognises that people's decision making on HE is complex and subject to multiple elements and influences. As far as possible your evaluations should seek to acknowledge this and aim to tease out what makes a difference to your target groups.

You should aim to be as precise as possible when identifying the measures you are using to capture the benefits of your interventions. For instance, the concept of 'aspiration raising' tends to be only weakly defined. Applying concepts such as this to your evaluation requires you to be critical and reflexive in order to make sure that the indicators you use capture well-defined changes and are relevant to your objectives.

Case study: Indicators for a sports coaching intervention at Loughborough University

Background: Loughborough University's SportMAD (Sport Making a Difference) project is designed to inspire young people from widening participation backgrounds to consider HE as a future option whilst raising understanding of what's required and developing transferable skills through sport and physical

activity.[3] Participating schools have been selected due to their higher proportion of widening participation countable students.[4]

How was the evaluation tackled: A range of indicators have been agreed supporting the evaluation, reflecting the specific aims of this intervention around increasing knowledge and understanding of sports-based HE provision. In addition to tracking progress to HE, the university implements pre- and post-evaluation surveys with individual participants. Indicators have been developed to determine:

Levels of understanding of what studying sport at HE level involves (pre- and post-participation)

Understanding of the qualifications needed to study sport at HE level (Do participants understand what's needed to progress to a selective institution?)

Attitudes to HE sports and exercise science (Are participants more or less likely to come to HE as a result?)

Psycho-social skills development (Has the project developed individuals' self-confidence?)

Result and learning: Consistent engagement and schools adherence to the evaluation was vital to the success of the approach. A key challenge was engaging with PE teachers who may not be familiar with evaluation processes (and may struggle with the process of completing consent and evaluation forms). Tackling this has required maintaining a consistent message as to the processes and data required.

How could this be developed further? The project is multi-faceted and it would be interesting to compare results for participants who engage to a greater or lesser extent. The intervention is being delivered in Year 10 and although the HE destinations of these students were being tracked it was not possible to say with certainty that any impact was due to the intervention. It would be interesting to have a control group or some other comparator to test whether the impact is associated with the outreach. Comparisons could be made with other coaching interventions if the intervention is not similar and the criteria used to target students for both interventions were the same.

3.4   Building in evaluation

Evaluation of outreach requires forward planning that is incorporated into the design of the outreach activities themselves. If the objectives and the indicators are not made clear when programmes/activities are at the design stage, the opportunity will be missed to ensure that methods and processes are put in place for the collection of the data relevant to the chosen indicators.

Your intervention theory helps you to be precise and purposeful in deciding what data to collect.

Tip: Evaluations should be planned alongside and embedded in outreach activity before its delivery. This helps to ensure data collection can go hand-in-hand and its quality is not compromised.

Over to you: Building evaluation into programme design
Is evaluation specified during the development of your interventions? (e.g. evaluation agreed in project specification, data collection mechanisms built in)

Suggestions for practice
Ideally you should aim to ensure that evaluation is in place from the start of activities, for example, by

---

[3] The work targets Year 10 GCSE PE students (aged 14-15), and is longitudinal involving a series of interventions: an initial session to introduce the coach and the SportMAD initiative; two 5 and 6 weeks of coaching delivered by trained Loughborough student sport coaches, taking place within local schools twice in the academic year (in November and February); a closing session to explore the skills gained as a result of the 5/6 week coaching programme; a celebration events for participation schools involving a sports tournament, a tour of halls, lunch in a catered hall of residence, and the opportunity to meet existing sports students to understand balancing playing sport with HE study.

[4] Schools identify the cohort they wish to participate in the project, under guidance from Loughborough's outreach team (determined with reference to Index of Multiple Deprivation (IMD), POLAR YPR Quintile of the school post code and National Collaborative Outreach Programme (NCOP) criteria). At least 66% of participants on the project should fall under a number of categories: White working-class boys; Pupils with learning difficulties; BAME; Young carers; Children in care; Refugees or asylum seekers.

agreeing the evaluation approach and action plan as part of the overall project specification.

Methods and processes for collecting date to inform your evaluation can be built in as part of your delivery approach rather than as an afterthought or additional activity. For instance, by setting out what data you will require from the stakeholders and participants on an ongoing basis, allowing time for data collection to take place, and making sure delivery staff are clear on what data needs to be collected and how.

Preparing a formal evaluation plan can be a good way of ensuring that evaluation is undertaken most effectively, i.e. by specifying what data will be required and when, and different people's roles and responsibilities.

Further help: Evaluation planning

Resource	Comments	Available at:
The RUFDATA tool	Framework for developing an evaluation (Reasons, Uses, Focus, Data, Audience, Timescale, Agency)	http://www.lancaster.ac.uk/fass/events/capacitybuilding/toolkit/planning.htm

Practice example: Using aggregate and individual-level data

If you are using quantitative methods, you want to aim for individual-pupil-level data rather than aggregated measures, unless there are a fairly large number of cases. If you only record changes across a whole group (e.g. the share of respondents) and are not able to match any pre/post responses you will not be able to say whether individuals benefited. Individual-level data allows you to track how each participant in your activity benefited (allowing you to infer whether your activity is likely to have made a difference to different individuals). Individual-level data is also important when tracking individuals across the student lifecycle. Plus, it is amenable to finer grained analysis, for example to show if there were any demographic sub-groups for whom the intervention was particularly beneficial or not.

Data based on looking at changes across a group of participants is less robust than calculating the change for each individual, and the sample would need to be large enough to ensure that the results were not skewed by any differences in the individuals included in the two groups being compared. [5]

Case Study: Brightside's ARCC (Access for Rural and Coastal Communities)

Background: ARCC used the internet to connect young people in schools in Kent, Sussex and the Isle of Wight with online mentors who acted as role models and provided personalised advice and support about university and career pathways.[6]

How the evaluation was tackled: An evaluation framework was developed before the intervention and put in place from the start. The evaluation involved conducting baseline and exit surveys for mentees, and an exit survey for mentors, as well as reviewing engagement data generated by the online mentoring platform. For Cohort 3, mentee and mentor surveys included questions from Brightside's quality and impact frameworks (which measure key outcomes and quality indicators of a mentoring relationship). This framework looks at six outcomes which Brightside believes are key enablers of young people making confident and informed decisions. Outcomes are divided into four behaviour outcomes and two capital outcomes and the evaluation framework is designed to measure each of these outcomes:[7]

> Behavioural Outcomes: self-efficacy, hope, growth mindset, coping

---

[5] For example it is more powerful to be able to say that x% of participants recorded an improvement rather than saying the proportion in the group changed from x% to y%.

[6] ARCC was delivered in 2016 across three 10-week cohorts. The project engaged 435 young people in 13 schools. ARCC Mentoring matched sixth form pupils with a mentor who could answer their questions about applying to and studying at university and/or working in a sector that interested them.

[7] The framework was based on outcomes identified during research for Brightside's Theory of Change and was developed with support from social investment consultancy CAN Invest. http://brightside.org.uk/what-we-do/theory-of-change/

> Capital Outcomes: human capital and social capital

The survey includes open-text, multiple-choice and rating-scale questions to collect both quantitative and qualitative data. As Brightside does not currently have a dedicated evaluation officer, the project manager undertook the data analysis.

Results and learning: The baseline and exit survey results showed that 45% of mentees increased their confidence in relation to achieving good GCSE grades; 82% of ARCC mentees were more aware of the range of further and higher education options available to them; 72% were more motivated to explore their future options; and 84% felt more confident in achieving their goals. Some unexpected challenges arose during the project with regards to meeting the evaluation objectives of a range of partners on the steering group. As a result of negotiation, the evaluation strategy was changed between the three different cohorts participating in the scheme, which made the results non-comparable. Moreover, the project collected aggregate data on how students and teachers viewed the scheme, and it was not possible to track the changes in views of individual students.

How could the approach be further developed? The strength of this case study was in basing the evaluation on a Theory of Change and integrating the evaluation activities into the project design from the start. The research design also integrated the views of teachers as well as mentees. A key weakness was not being able to track the changes at individual level, and therefore, currently ongoing evaluations are using individual-level data to track the performance of individual students on the scheme. In addition Brightside is joining HEAT (Higher Education Access Tracker) to implement longitudinal tracking of mentees in order to gain greater insight into its long-term impact. Resources for evaluation can be a challenge for third-sector partners and it might be that a sustainable evaluation needs to be part of the project package funders agree to from the start.

4   Evaluation Design

4.1   Which Evaluation Type should I aim for?

Deciding on the best evaluation design can be tricky, and a range of factors will come into play, including the purpose of the evaluation, the nature of the outreach intervention, your evaluation capabilities and the availability of existing evidence of impact. Some types of evidence are more robust or trustworthy than others. The objective of working to standards is to help to assess the level of confidence that you have in the evidence in terms of showing that your outreach intervention is having the outcomes and impact you desire.

A range of factors will determine which type of evaluation is most appropriate in different circumstances.

There is no hard and fast rule about which type of evaluation applies to which type of outreach activity. However Type 1 is a minimum requirement for all interventions to provide a clear articulation of why the intervention is necessary and a good idea.

Type 1 evaluation is a minimum requirement.

The intensity and cost of the outreach will play a part in deciding which type of evaluation is proportionate to the intervention: the greater the cost and intensity the more confident we need to be that it is making a difference. So for a long-term or multi-activity interventions, a summer school or other HE residential programme, or mentoring and coaching programmes, a Type 2 or 3 evaluation would be expected.

More intensive and costly interventions usually require higher standards of evidence.

The availability and strength of the existing evidence base which already exists to evidence the results for your or comparable interventions also has a bearing. Where there is no or only weak evidence, there will be most need to generate robust evidence to show the impact of it. Other considerations include the specific research questions your evaluation seeks to address.

Evaluation should aim to fill gaps in the existing evidence.

Different types of evaluation are not hierarchical. The type of evaluation you should undertake depends on what you want to be able to claim from your evaluation findings. As a general rule Type 1 evaluations help you to present a plausible rationale for why you are doing what you do. Type 2 evaluation is important where you need to report evidence that those receiving an intervention treatment have better outcomes where this is uncertain, debated or needs more investigation. This type of evaluation can demonstrate whether or not your activity is worthwhile or not to continue (without establishing definitive direct causal effects). Type 3 evaluations are important to use if you think an intervention is going to be effective but you need to know for sure if it works and need to be confident in the result (e.g. before rolling it out further). N.B. if you can already show that something is going to provide the benefit you desire in a particular context then you probably don't need to go to the expense of an experimental trial. You should aim for the best evidence within each type of evaluation. The sections below are designed to help you to strengthen the quality and reliability of your evidence.

You need to choose the type of evaluation that best supports the claims you want to make.

More costly and intensive activities generally require a more robust evaluation design.

	Description	Evidence used	Claims you can make
Type 1: Narrative	The evaluation provides a narrative and a coherent theory of change to motivate its selection of outreach activities in the context of a coherent outreach strategy	Evidence of impact elsewhere and/or in the research literature on outreach effectiveness or from your existing evaluation results	We have a coherent explanation of what we do and why Our claims are research-based
Type 2: Empirical Research	The evaluation collects data on outcomes and impact and reports evidence that those receiving an intervention have better results, though this does not establish any direct causal effect	Quantitative and/or qualitative evidence of a pre/post treatment change or a treatment/non-treatment difference	We can demonstrate that our interventions are associated with beneficial results.
Type 3: Causality	The evaluation methodology provides evidence of a causal effect of an intervention	Quantitative and/or qualitative evidence of a pre/post treatment change on a treated group relative to an appropriate control or comparison group using an appropriate and robust research design	We believe our intervention causes improvement and can demonstrate the difference using a control or comparison group

Tip: You should select the most appropriate methodology to your context, the objectives, and any practical constraints. The aim should be to work towards increasingly robust evidence of impact.

Over to you: What is your aim in doing evaluation? Who is the audience?
What are your goals as a higher education provider or third-sector organisation in doing evaluation? What are your questions/what is it you want to know? How will you use the evaluation as part of your internal outreach development? How can the impact evaluation help feed into statutory returns?

4.2 Type 1: Narrative evaluation

Type 1 is a requirement of all outreach provision – i.e. you can provide a narrative to explain the selection of outreach activities in the context of a coherent outreach strategy. There is no hard and fast rule as to the form of this narrative, rather it should reflect an evaluation approach appropriate to your particular context. It might be based, for example, on articulation of a clear Theory of Change (see above) or an evidence-base for the activities being undertaken, either referring your own existing evidence of impact and/or in the research literature on outreach effectiveness.

All types of outreach activity should be underpinned by a narrative on why they are being delivered in this way.

A search for existing evidence could seek to identify what ideas have previously been tried to address the issue, or could focus on a particular type of evidence to see if there are previous studies that have been conducted to test its effectiveness. A thorough search might also include an examination of whether the intervention has been studied regarding its effects in relation to other issues or target groups as well.

The narrative can be informed by your own or other people's evidence.

The process of checking the existing evidence of the effectiveness of interventions relating to the issue that you are aiming to address will involve collecting evidence of the effectiveness of interventions that are relevant to the issue and considering the overall picture that emerges about an intervention if multiple studies exist. A systematic review is a well-known approach designed to summarise existing evidence (a type of literature review that uses systematic methods to critically appraise and synthesise studies). This technique would be most useful where there is a substantive research question to be addressed and several empirical studies have been published.

A review of existing evaluations can become a source of evidence (e.g. a systematic review).

If conducted on a systematic and rigorous basis, a review of the evidence becomes a powerful and compelling type of evidence in its own right, and even when conducted on a 'light-touch basis', a completed evidence review should inform decisions on the programme. However, a key issue in the context of widening participation outreach is the lack of credible evidence about the impact of different types of interventions. Therefore it may be necessary to take a broad approach to potential sources of evidence.

There is a general lack of evidence on impact of outreach activities.

At the same time it is important to think critically about the validity and relevance of the evidence. Questions to ask include: whether there is convincing evidence of an impact on outcomes that are of interest; whether any correlation observed between the activity and outcome was down to a causal relationship; and whether the findings are transferable to your particular context. Factors such as the type of evaluation study, the number of participants, and the quality of the research and analyses will affect the level of confidence you can have in the results.

The conclusions you can draw depend on the quality of the evidence.

Tip: Identifying that there is little or no evidence for the activities, or only uncertain evidence on the results, would establish that it is appropriate to conduct a programme evaluation study: the guidance provided below for Types 2 and 3 should then prove helpful.

Over to you: Evidencing a Type 1 evaluation
Do your programmes have a Theory of Change or logical framework that demonstrates an understanding about what works in what context? Can you point to evidence to support the processes identified in programme activities?

4.2.1 Strengthening Type 1 evidence

Unlike the types of evaluation, which are not hierarchical, there is a **hierarchy of evidence**. It is thus possible for providers and OfS to judge the claims a provider makes with regards to having a Type 1 evaluation.

Strong evidence is grounded in the research, underpinned by a rationale and a coherent approach.

The following table summarises what is weaker and what is stronger evidence as part of a Type 1 evaluation, including best evidence.

Dimension	Weak evidence	Developing Evidence	Example of best evidence
Underpinning rationale	There is no intervention theory or rationale for the activities. Individuals work in silos in an ad-hoc fashion. Only those undertaking evaluation work know some of the emerging Theory of Change.	Appreciation of the context of access and participation work. A intervention theory exists that is underpinned by the literature. The intervention theory is shared among widening participation	The intervention theory is documented, available and shared, within and beyond widening participation teams across the whole institution, e.g. through informal practices, meetings, seminars, committees.

Dimension	Weak evidence	Developing Evidence	Example of best evidence
	Retro-fitting of the intervention theory to fit the activities or evaluation.	practitioners involved (e.g. through a shared Theory of Change).	Evidence of shared knowledge and opportunities for a range of stakeholders to comment and contribute to the intervention theory Development of specific programme theories for each activity.
Grounding	No engagement with literature or current debates. Ad-hoc development of activities and evaluation. Retro-fitting of purpose of the activities or evaluation.	Critical engagement with and reflection on the literature. A clear link between the intervention theory and current and planned institutional practice and evaluation.	Existence of a systematic, up-to-date review of relevant literature, including theoretical, empirical, and policy literature with full references. The critical engagement with and reflection on the literature is accompanied by suggestions for further research; and/or actual additional research being undertaken or commissioned. Evidence of appreciation of the national, regional context as well as institutional context within which the activity operates. Clear evidence of how the literature review is used to inform the intervention theory.
Engagement	No evidence of engagement in debates beyond the institution.	Some participation as delegates or network-mail users in conversations beyond the institution, but no evidence of contributing to those debates or these debates feeding back to enhance practice. Engagement as participant and contributor with cross-institutional, regional and/or national debates through professional networks and conferences. Evidence of linking wider debates to institutional practices and policies.	Presentation at cross-institutional and/or national (or international) conferences. Creation of shared resources practitioners in the sector can use. Evidence of how participation in cross-institutional and national conversations feeds back into institutional practices and/or policies.
Criticality	No re-evaluation of the literature and context of the activity.	Awareness that contexts might change but no structured way of feeding this back into policies and practice. Review cycles exist and are planned within the regular work-cycle.	Set review cycles that re-consider whether the intervention theory and related practices need updating in light of developing policy and evidence/theory development contexts. Results are documented, available and shared, and reflected on, e.g. through informal practices, meetings, seminars, committees.
Coherence	Activities and evaluation are developed in isolation.	There is some connection between activities. Most activities are joined up and underpinned by evidence.	A joined-up, holistic approach which considers the relationship between activities and cohorts and looks for opportunities to connect these.

4.3   Type 2: Empirical research

It is possible to choose from a range of methodologies, and some evaluations use a combination of methods. Your choice of method should be guided by the questions you want to answer. You should think about the purpose of evaluation and then select a design that will enable you to achieve your aims for the evaluation. It may not always be possible to choose the strongest evaluation design as a lot would depend on the scope of the study and the data that is or will be available. Annex 1 provides guidance on how different types of impact evaluation might be applied to different types of outreach.

A range of methods can be used, depending on the opportunities and constraints for evaluation.

Broadly speaking, **quantitative methods** are useful if you are primarily concerned with evaluating the scale of the outcomes, for example, the number of participants in your outreach programme who went on to achieve 5+ A*-C in GCSEs or were offered a place in a highly selective university. These can also be used for estimating short and long-term outcomes for participants, gauging effectiveness of interventions in improving for example, attainment, participation, and HE access of participants.

Quantitative methods capture the numbers of outcomes achieved.

If you only currently collect data at the end of an activity (for example, through an end-of-event questionnaire), consider collecting some data from participants before or at the beginning of the event so that you have a point of comparison. If you are already collecting before and after data from your participants, you could consider collecting data again some period later (maybe three to six months) to see whether any changes in knowledge, attitude or behaviour have remained. The data could be limited to a single question, or include different aspects, depending on the objectives of your intervention and what outcome(s) you seek to affect.

Consider collecting evidence at different points in time.

Tip: Any pre- and post-intervention data should be available for a reasonable number of participants (at least 30) and if you use a sample you should demonstrate representativeness with your participants as a whole

Quantitative data from outreach can be analysed in different ways. ***Descriptive statistics*** provide simple summaries about the participants/groups. Together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data. ***Inferential statistics*** are useful when you want to compare two sets of data on a single measurable outcome to see if there is a difference (for example, results before and after the intervention, between groups of participants at a point in time, or between groups over time (difference-in-difference). As an example, for an intervention aimed at improving understanding of maths amongst Key Stage 4 pupils, you might want to know whether the intervention group differs on the outcome measure from a control group. To do this you will analyse whether intervention and control groups differ in GCSE maths test scores. You may also want to see by how many points the overall scores improved and you will then consider other parameters such as prior attainment in Key Stage 3.

Descriptive statistics summarise the results in relation to your specific participants whilst inferential statistics attempt to show how the results might be generalised (i.e. the results which might be expected in target cohorts in general).

You can match people on the basis of similar characteristics (e.g. socio-demographic criteria), or in relation to their prior achievement level (i.e. performance in standard tests or in any test you are using to measure change). An alternative simplistic approach might be to compare against other groups you know about – for example, if you have changed what you deliver you could compare against the group in a previous cohort (year group) who did not receive the new intervention, or a cohort in a non-intervention school. However, there are likely to be differences between the two groups, which you will need to take into account when interpreting your results. This approach is not very robust but could be a starting point for considering more systematic comparative studies in future.

Using a comparison group is a good way of establishing the outcomes associated with your outreach by showing what might have happened without the intervention.

You can estimate a range of effect sizes such as attainment gap, odds ratio or risk ratio to assess whether the intervention has brought in a change and whether the change is positive or negative. If you are testing a hypothesis, you could declare that you are looking for x outcome in advance and then test whether this has been achieved based on conventions around the significance of p values.[8]

Statistical techniques can be used to show the strength of any observed effect on outcomes.

Further help: Using statistics		
Resource	**Comments**	**Available at:**
Web Center for Social Research Methods	Provides a discussion of social research methods, including guidance on descriptive and inferential statistics	https://socialresearchmethods.net/kb/statdesc.php

Qualitative research usually aims to understand people's beliefs, experiences, attitudes, behaviour and interactions. Qualitative evaluations are useful to gain insight into why something happened (e.g. perspectives of the processes involved in achieving the desired result). There are a variety of tools such as essays, interviews, focus groups, scenarios, projects, case studies, artefacts, capturing personal experiences, introspection, visual texts, portfolios, direct observation, role play or simulation.

Qualitative methods capture processes and perceptions of the changes associated with an activity.

Just as one of the benefits of qualitative data is its richness and variety, so one limitation is a tendency to create selective and narrative accounts of the data. It is all too easy to cherry-pick through your qualitative data to find responses that match what you want to claim. It is therefore important to consider robust qualitative data analysis. *A deductive approach* to qualitative data analysis is 'top-down' – where you have predetermined what categories you are going to look for (this approach is best when you are searching for something very specific). On the other hand *an inductive approach* is 'bottom up' – where you read your data closely and create codes/categories that seem best to describe what the data is revealing (this is best when you are less sure about what you might find or what the 'top-down' themes might be).

It is crucial that your approach to qualitative data analysis is systematic and unbiased.

Skills you will need: Collecting and analysing qualitative data, awareness of ethical implications and, as applicable, ethical consent processes.

Skills you might need: Being DBS-checked, specific skills in developing scripts for focus groups or interviews, skills in observation and/or development and analysis of open-ended questions.

Further help: Qualitative data analysis		
Resource	**Comments**	**Available from:**
Braun, V. and Clarke, V. (2006) Using thematic analysis in psychology.	Very helpful in outlining how you might use inductive analysis to analyse qualitative data.	Qualitative Research in Psychology, 3 (2). pp. 77-101. ISSN 1478-0887 http://eprints.uwe.ac.uk/11735

---

[8] A p value shows the result of statistical significance testing. For example, a p value of 0.05 is used to suggest that the results are statistically significant (denoting a 5% chance that the results were down to chance).

Resource	Comments	Available from:
Qualitative Data Analysis Guidance	Online resources which support the use of qualitative methods	https://www.achievability.co.uk/evasys/how-to-effectively-carry-out-a-qualitative-data-analysis
How to Analyse Qualitative Data	A useful site if you want to know how to analyse qualitative data. Includes sections on 'How to conduct interviews' or 'How to conduct participant observation'.	http://www.emeraldgrouppublishing.com/research/guides/methods/qualitative.htm

Practice Example: Qualitative evaluation looking at development of HE-related skills and attributes

There are many different kinds of qualitative data and you could create your own method of data collection, if you feel a particular approach will give you the data you need. Attention needs to be paid to the design of your data collection so that you get the best data possible: it is easy for interviews or focus groups, for example, to generate lots of data but you would need to use focused methods in order to make sure you address your key evaluation questions. All data can be collected and analysed badly, and used without rigour to make claims, but perhaps this is especially true of qualitative data, where it is very easy to collect data unsystematically, and to over-rely on 'vox pop' types of feedback. It is usually helpful to combine qualitative methods with 'hard' data for example on progression outcomes.

Case study: Royal Northern College of Music (RNCM) Young Company

Background: Young Company offers performance training delivered by leading professionals such as directors, choreographers and vocal coaches working with RNCM students. The project encourages the group to consider HE and rehearsing in an HEI provides opportunities to meet with HE students.[9]

How tackled: The project is working to develop a range of skills and outcomes and is evaluated in several ways including pre- and post-project questionnaires, use of video recording, talking and listening; and observation. By keeping in regular contact with the participants, the outreach manager is able to track individual participants' HE outcomes over time.

Results and learning: Talking to the participants and listening to their views is central to the evaluation approach and the outreach manager considers the most valuable information is gathered in this way. From the outset participants are encouraged to talk to staff who work to develop a relaxed atmosphere conducive to this (for example, by asking for suggestions for careers talks). Observation is also important: through watching and analysing the skills of the group a clear picture is built up of where they are in terms of skills development and adjustments made accordingly. The young people are assessed by looking at the skills level when they join the project and how they improve at various stages of the process; the provision is tailored accordingly through small group and individual work during project sessions, and giving them exercises to do outside of group time. Through working closely with the group and staying in contact over time, tracking progression is very easy. The majority of the groups have moved on to HE, and recent destinations include a wide range of specialist arts education providers and leading universities.

How could this be developed further? The developing focus on pre- and post-intervention measures could be strengthened with a view to developing a greater understanding of the counter-factual, i.e. what would have happened had the intervention not taken place.

Mixed-methods research utilise both qualitative and quantitative approaches. A mixed-methods approach can overcome the limitations associated with any single evaluation design, whilst also offering opportunities to explore and interpret the work and to address

Mixed-methods evaluations use qualitative and

---

[9] Cohorts of around 35 people take part and receive training in singing, acting, improvisation, and dance. Performances are staged in a professional venue and supported by experienced production crew, and a live orchestra. The group receives training in related skills including stage combat and acting for film and television. In addition, sessions are run on applying to performing arts institutions, preparing for auditions, controlling audition nerves and choosing where to study. Plus there are careers talks, and visits to performances in other venues. At least five student mentors are attached to the group each year.

a question at different levels. Although you will need to make sure that sufficient time is available to ensure the research is systematic and credible, and that you explore and resolve any difference between findings from different types of data.

Qualitative and quantitative data can be highly complementary: very often the quantitative data will tell you *what* happened, whilst the qualitative data will tell you *how* or *why* it happened. A well-chosen qualitative data set, student interviews for example, might be important to explain the effectiveness of the intervention. If you are only collecting data from young people through questionnaires, consider undertaking focus groups or group interviews with a sample after a period of time has elapsed – this will give them the opportunity to reflect on their experiences and add richness to the questionnaire results. If the quantitative data showed overall success, but that for some groups it was less successful, then these interviews might tease out the reasons for this. Using an 'authentic task' exercise can provide additional observational data that can complement self-reported data (which only collects participants' subjective perspective on the aspect under investigation). Interpretation of data can progress in stages – i.e. early results from qualitative research can influence future stages in the research process (e.g. focus groups informing questions for a quantitative survey which could help to generalise, to a degree, the qualitative data).

Mixed-methods research also has the advantage of drawing in data to reflect a range of perspectives. If currently rely on gathering evaluation data only from participants, consider triangulating the results by gathering data from adults working with the targeted young people, including parents and teachers. To develop this practice further you might consider using short telephone interviews – many will prefer this (response rates will be stronger) and you will collect richer data in a more robust way than using questionnaires.

Some common evaluation methodologies are based on the use of mixed-methods research or triangulation of data from different sources. Evaluation methodologies including 'realistic evaluation' and 'contribution analysis' are designed to deal with the problem of attributing outcomes and impact to an intervention when working in complex systems in the sense that they involve specification of how activities will lead to changes and identification of the contextual factors that may affect them.

quantitative data in a complementary way.

Qualitative data can help in interpreting quantitative results or identify the processes involved. To an extent, quantitative data can help to generalise up from qualitative conclusions.

Different methods can help to collect evidence from a range of perspectives.

Building the case for contribution strengthens the argument in favour of an intervention's impact.

Tip: Avoid seeing quantitative data as better than qualitative data: they often do different things, and your choice of what data you collect should be based on what you need to know.

Over to you: Selecting a research method
Which research method is most appropriate given the size of your cohort(s) and your access to participants/controls and data collection potential? Can you identify participants and controls? What types of data will you need to collect and how? What analytical strategy will you use to generate results once you have collected the data? Who do you need to work with?

4.3.1   Strengthening a Type 2 evaluation

Aspects of good practice for a Type 2 evaluation encompass those noted above as part of a Type 1 plus additional considerations.

Different practices are associated with weaker and stronger evaluation evidence. The following table summarises what is weaker and what is stronger evidence for a Type 2 evaluation, including examples of what providers can do to generate the best evidence.

It may not always be possible to choose the strongest evaluation design as a lot would depend on the data that is or will be available. However whatever your starting point it is important to think about how to put in place actions to strengthen what evaluation evidence you are collecting over time.

Strong evidence uses appropriate indicators, uses valid tools, robust sampling, appropriate analysis and

Dimension	Weak evidence	Developing Evidence	Example of best evidence
Indicators of your impact	No concept of measuring the success of your activities against indicators of impact.	Outcome measures identified which directly relate to the aims of your activities and the impact you want to make.	Identified outcome measures in the short, medium and longer term, and mechanisms for the collection of data which provides evidence that allows you to report a change in at least one of your project's key outcomes (although not establishing any causal effect).
Data collection tools	Data not related to the aims of the intervention. Information not systematically collected. There is only weak engagement with how measures of change are defined.	The measures of change are precisely defined and there is clarity on their relevance to the intervention. Validation of any data collection tools through cognitive testing and piloting.	Critical and reflexive engagement with the measures of change. Use of data collection tools that have been pre-validated. Use of systematically collected and reliable administrative data.
Data collection	Data collected at one point of time only. Retrospective data collected at the end of the project (i.e. no proper baseline measurement). Only collecting data from one source (e.g. outreach participants).	Pre/post data (minimum two points in time). Gathering feedback data from adults involved with the targeted participants (parents and teachers).	Pre/post data collected (and preferably during the intervention). Use of the same data collection method with the control and participant group over the same period of time. Application of a systematic mixed-methods design which triangulates results from multiple perspectives.
Making comparisons	Study does not make a convincing comparison with what might otherwise have happened without the outreach. For example, the use of a comparison with the previous year's cohort.	Study uses comparators but there is no attempt to control for individual factors. Study has a comparison group, but it may exhibit some weaknesses, e.g. the groups lack comparability on key participant variables.	Study design uses a convincing comparison to identify the impacts e.g. matched comparators based on key variables.
Sampling	Use of only a small number of cases.	Evaluation across a larger sample.	Evaluation of your whole cohort (where the numbers permit). A sufficiently large or well-selected sample that enables you to be fairly confident that any observed improvement was down to your project.
Outcome measures	Reliance on self-reported data through questionnaires.	Use of objective measures (such as applications to HE). Gathering perspectives on the impact on outcomes from participants and other stakeholders working with the young people (teachers).	Tracking participants over time into HE. Use of systematic qualitative data collection (e.g. interviews, focus groups) to gain in-depth insight.

Dimension	Weak evidence	Developing Evidence	Example of best evidence
Analytical strategy	Use of descriptive statistics only (simple percentages) to measure changes.	Use of inferential testing to determine whether the changes can be ascribed to the activity rather than to chance.	Use of multivariate analysis that takes account of background variables (e.g. gender).
Addressing study limitations	No acknowledgement of the limitations.	Recognition of the likely limitations of the approach and issues (e.g. selection bias*).	Recognition of the likely limitations of the approach and putting in place attempts to mitigate these (e.g. controlling for selection variables).

* The issue of selection bias is that the observed differences could be due to how the intervention group is selected, for example if they are more motivated, have a pre-disposition for HE or high levels of prior attainment.

Over to you: Measuring change

Do you measure the changes associated with your interventions against a counter-factual, i.e. compared to what might have happened otherwise had the interventions not been in place? Do you collect evaluation data at different points (before and after (and preferably during) participation in outreach)? Does your research design involve use of comparison groups? Do you use inferential statistics, where appropriate, as well as descriptive statistics?

Further help: Common evaluation methodologies

Resource	Comments	Available at:
Pawson, R. and Tilley, N. (1997). Realistic Evaluation. London: Sage.	Seeks to understand causal mechanisms by drawing on different forms of data, using a model that considers the context, mechanisms for change, and desired outcomes.	An introduction to realist evaluation including a downloadable chapter from Pawson and Tilley (1997) www.communitymatters.com.au/gpage1.html
Yin, R. K. (2018) Case Study Research and Applications: Design and Methods (sixth edition). Thousand Oaks, CA: Sage	Assumes that results are affected by the physical and human context and that a holistic approach is required. Emphasises triangulation of methods and perspective to test the underpinning intervention theory.	SAGE Publications Inc

Practice Example: Mixing Methods

In building a case for whether or not an intervention is having an impact, it is often useful to use 'triangulation' of data – that is the use of the findings from different data analyses (drawing on a range of perspectives and drawing on both qualitative and quantitative evaluation methods). Drawing on a range of sources of evaluation evidence can also help to identify and explain the conditions under which the theory of change is seen to operate. A narrow focus on impact evaluation can be limiting in terms of helping to understand why and how the outreach is most effective. For example, regression-based analysis of data obtained from randomised control trials (RCTs) might be able to provide an explanation of how an observed impact varies across students, but is likely to be limited in explaining what made the difference on the ground and under what conditions.

Case Study: Sutton Trust's 2018 UK Summer School programme

Background: The project is a residential programme held at a leading university in the UK.[10]

---

[10] There are 11 universities and over 100 course combinations to choose from. Some 2,000 places are available and for the 2017 programme 11,000+ applications were received. High achieving Year 12 students who have always attended a state-funded school or college (non-fee paying) in the UK meeting additional eligibility criteria are selected. List of eligibility criteria available here: https://summerschools.suttontrust.com/eligibility/. Sutton Trust manages the application system and reports on the eligibility of applicants, and university partners report on programme delivery. Sutton Trust also oversees delivery of outreach activities at all 11 universities. These universities report annually on programme outcomes and impact. Targeting and impact reports are produced each

How the evaluation was tackled: The Sutton Trust case study involves using an external evaluator to complete the evaluation, along with the use of a common evaluation framework across the organisation. Two forms of evaluation are taking place for the summer schools cohorts:

- Process evaluation – conducted by the National Foundation for Educational Research (NFER)
- Impact evaluation – conducted by Educational Research and Analysis (ERA). The impact evaluation will look at changes in attitudes and aspirations using a baseline and exit survey and then look at destination data of students who take part in the programme. The approach is Type 2 evaluation practice: data is collected on pre- and post-programme attitudes and student destinations are tracked (more comprehensively from 2017 onwards).

Baseline and Exit surveys: Sutton Trust baseline surveys are delivered prior to a programme starting. Exit surveys are delivered at the end of the programme.

Core questions: All post-16 programmes use core questions, which run across programmes. This is mirrored in the pre-16 programmes. Each of the core questions is linked to at least one evaluation outcome. Likert scales are used to capture participants' perceptions and analyse changes in attitudes numerically. Post-16 programmes use a nine-point Likert scale and pre-16 programmes use a five-point Likert scale. This is to simplify surveys for younger students and allow for a more granular analysis of post-16 programme results.

Programme specific questions: Surveys may also have programme specific questions as part of the survey. These are developed by the programmes team, in conjunction with evaluators. This allows the trust to tailor surveys to specific programme needs and outcomes

Analysis: Evaluators are asked to measure changes in aspiration and attitude against evaluation outcomes. Wherever possible, this is through matching an individual's response to the baseline and exit surveys. If this is not possible, cohort averages are used.

Results and learning: Effective evaluation relies on responses to surveys. The trust works very closely with young people and delivery partners to ensure a high completion rate can be obtained. The evaluation is longitudinal, as it will look at destination data of the cohort as well as immediate impact through surveys (uses a control group). It will, however, only cover one cohort. Sutton Trust plan to conduct a five-year longitudinal study of destination data (last completed in 2016).

How could the approach be further developed? The Sutton Trust is looking to strengthen evaluation methods, however, there are challenges as they do not run the programmes directly since the provision is held at university partners. The programme overall therefore includes a lot of different variables – such as selection criteria, university attended, course attended – which makes it difficult to create robust control groups. Also, as the programme is well established and has been running for 20 years it is challenging to run a 'test' evaluation such as an RCT. The trust is exploring how a Type 3 evaluation might be used with the programmes in practice, without impacting the ethics of the programme itself.

Key terms: Measuring the difference in outcomes

Difference over time: Measuring the difference can involve looking at participants and making a comparison of results before and after participation e.g. changes in attainment and/or aspiration measures. Difference-in-difference evaluation does both, i.e. this method looks at a before-and-after change in participants relative to that of non-participants.

Comparison group: This group should be generally comparable to your participants, and is used to demonstrate how results differ to those taking part in the activity. Usually comparison is between participants and non-participants, although there could be more than two groups (e.g. groups receiving different types of interventions).

Control group: A control group is similar to a comparison group but selected using a scientific approach to ensure that the group of participants and non-participants are as similar as possible by demographic variables (i.e. distinguished as far as possible only by participation/non-participation in the intervention).

---

year and these reports are then shared with the trust's senior management, trustees and development board. The evaluation partner provides an independent impact report. External reports are provided to programme donors.

Internal and external comparison groups: An internal comparison group would be made up of people you know who are similar to participants (e.g. if you have a competitive application process for your intervention because places are limited, you could use the group of rejected applicants as a comparison group (with their consent)). For an external comparison group, people are identified who have not been identified through your activities (e.g. other pupils in schools who were non-participants).

4.4 Type 3: Establishing causality

Using comparison or control groups is important where the evaluation is seeking to explain outcomes for some young people compared to a cohort (e.g. those in a school cohort who took part in an outreach programme, compared to those that did not), or compared to the wider population (e.g. similar young people in the population who did not receive the intervention). A Type 3 evaluation involves a methodology that is capable of providing evidence of a causal effect of an intervention. Type 3 evaluations give more confidence than Type 2 because they utilise more robust methodologies.

Type 3 evaluations show the effects whilst controlling for alternative explanations for outcomes.

An experimental design eliminates factors that influence outcome except for the intervention being studied by random assignment of participants and control of the study including use of control groups. A quasi-experimental design is used when randomisation is not possible and other techniques are used to build a comparison group that is as similar as possible to the intervention group in terms of their pre-intervention characteristics and conditions.

Often one of the most challenging tasks is to identify an appropriate comparison group. Different experimental and quasi-experimental methods are described in the table below. There are various comparisons you can make to test the efficacy of interventions such as univariate, bivariate and multivariate analysis, cross-tabulations and regression analysis.

Different research designs are possible.

Type 3 evaluation research designs

Design	Comments	Issues emerging for use in an outreach context	Situations where this approach might be appropriate
Randomised control trial (RCT) (the groups are allocated randomly)	Avoids selection bias issues.\ Takes account of the likelihood of changes in participants over time and regression to the mean effects.\ Capable of generating statistically powerful results.	This is sometimes seen as a deficit model by critics, however, if there is enough of a question about an intervention's efficacy to warrant conducting an RCT, those who 'miss out' on the intervention can be compensated by other forms of support at another time.\ There can be issues around ensuring the control group is not exposed to any outreach intervention or benefits in other ways (which would contaminate the results).\ RCTs are not appropriate when the outcomes of interest are far in the future.	Could be suitable for evaluations of 'light-touch' interventions where the activity can be isolated.\ Most appropriate where the theory underpinning why an intervention works needs to be tested.\ Usually requires targeting criteria that are relatively general (meaning there is no reason why one young person was chosen to participate while similar ones were not).
Propensity Score Matching (PSM) (compares outreach group with individuals who did not receive the intervention but who were as likely to have done so)	A mechanism for retrospectively assigning a comparison group (thus avoiding the issues involved in randomisation).	Relies on access to an appropriate large-scale dataset which contains appropriate fields to assess eligibility (e.g. pupil census data).\ Some data intermediaries have in place mechanisms for selecting a comparison group (e.g. HEAT, UCAS STROBE).	Could be suitable in situations where selection of the intervention and control group on a random basis is not feasible.

Design	Comments	Issues emerging for use in an outreach context	Situations where this approach might be appropriate
Regression discontinuity design (RDD) (compares the outreach group with individuals who are essentially equivalent but just fail to meet the criteria for participation on one key characteristic (e.g. family income)).	Requires a relatively large number of cases to be included. Assumes a linear relationship between pre- and post- testing.	Only appropriate for interventions where there is very clear selection criteria and access to a pool of suitable candidates who were not selected to take part.	Could be suitable for evaluations of summer schools or academic support interventions where there is an application process that is over-subscribed.
Matched research designs (involves development of a comparison group which mirrors the intervention group)	Possible to undertake with relatively small numbers of participants. Assumes the treatment and control group have the same characteristics.[(1)] Controls could be internal or external.[(2)]	If you have access to learner data in schools you can match by demographic characteristics or for example take the next pupil above and below on the school register (taking sex into account) whose parents consent to inclusion.	Situations where there are close relationships in place with stakeholders and data sharing protocols are in place. However there can be contamination issues if the comparison group benefit from the outreach.
'Natural experiment' (where the experimental and control groups are selected by factors which are not controlled)	Situations where setting up a controlled experiment is difficult or unethical.	Relies on access to an appropriate large-scale dataset, for example, where the experimental and control groups are chosen from different points in time before and after a new activity is implemented, or those in a school who have been part of an outreach programme, compared to those who have not.	Where data is available for comparison and the external conditions and characteristics of the groups are stable enough to infer that any difference in the observed result is due to the intervention.

(1) This is the main weakness of the matched approach as it not possible to discount the possibility that any difference observed was down to a variable that was not measured. (2) Where there are only a small number of participants it may be feasible to match all of them (e.g. 20 participants with 20 controls with the same characteristics). For larger samples it is possible to have few controls (the case:control ratio should 3:1 to maintain the power of the test and the more variables, the higher the control:case ratio needed).

Using comparison or control groups is important where impact evaluation seeks to explain outcomes for some young people compared to a cohort (e.g. those in a school cohort who took part on an outreach programme, compared to those that did not), or compared to the wider population (e.g. similar young people in the population who did not receive the intervention). Experimental research methods – such as an RCT where the participants and controls are selected at random – are very strong evaluation designs. However, it is important to consider what works best for your outreach activity. For example, it may not be possible to have an experimental set-up with 'laboratory conditions' (where participation is controlled and limited). Furthermore, RCTs can sometimes be harder to implement for long-term social interventions than for more clearly defined short-term interventions. They tend to be most appropriate where the intervention is quite simple and relatively binary. Quasi-experimental designs aim to overcome this problem by using administrative datasets in combination with data collected from participants through the evaluation.[11]

Not all types of outreach will lend themselves to experimental designs.

---

[11] One approach is to match people on the basis of similar characteristics (e.g. gender, ethnicity, free school meals (FSM) status) and in relation to prior achievement level (i.e. performance in standard tests or in any test you are using to measure change). Entering their details into two lists and ranking them in order would allow you to 'pair' those who are closest on the characteristics you are interested in. The 'pairs' would be split across the intervention and control group. After the outreach has been completed, you would compare outcomes or performance in a post-test measure to generate findings on the extent to which the intervention made a difference (for example by looking at the average difference between the people in the two groups).

Type 3 evaluations are challenging because they demand an appropriate design, dedicated data collection from different groups, mechanisms for ensuring that appropriate data can be obtained and the appropriate consents to use the data put in place and a robust analysis.

Type 3 evaluations are challenging.

Selection bias is a major problem. Outreach providers use a range of criteria for targeting and the characteristics of those involved can differ, especially as for some activities the decision about who takes part is made externally (e.g. in schools and colleges). It is important to realise there can be various differences between these pupils as their personal backgrounds and schools might differ. Allowing within the analytical framework for undertaking analysis of different sub-groups is therefore useful in order to see which groups get the best results (and ideally under what conditions). You will need to think about the extent to which your design compensates for any potential selection bias by ensuring the comparison/control group is as close as possible to the group taking part in the activity. This issue of avoiding selection bias is particularly important in the context of delivery of outreach given that those who participate in an intervention may already be well motivated or with an HE-orientation. This is especially the case with targeting – how to avoid a self-fulfilling prophecy whereby the people chosen are already predisposed to certain behaviours. Ideally your data analysis strategy should control for things like the level of motivation or prior attainment. However, this can prove tricky in practice because there is no easily accessible measure for these types of influencing factors.

Selection bias is a particular problem in the context of outreach because it can be hard to take account of the characteristics and other influences on participants.

There is potential for 'contamination' effects when working in a natural environment such as a school or college. If individuals in the comparator group have had some kind of intervention the efficacy of the activity being evaluated will show a 'dimmed effect'. Thus, the comparator needs to be chosen very carefully when isolating and assessing the contribution of your intervention and any known limitations need to be acknowledged.

Ensure the comparison group has not received HE access support.

It has been argued that it is not always fair to withhold an intervention from some widening participation students and not others, however, given that the evaluation is needed before we can actually show that an outreach intervention is beneficial to their progress, the counter argument contends that it is unethical not to attempt to establish the intervention's effectiveness (especially if there is scope to roll it out further in future). In some cases comparisons can be made without withholding interventions – for example when comparing applicants and non-applicants to an activity or looking at how one form of delivery compares with another.

There are ethical arguments for and against using control groups.

Skills you will need: Research design for planning, good working knowledge of statistics for undertaking analysis, awareness of ethical implications and, as applicable ethical consent processes, expert use of software for statistical analysis

Skills you might need: Data linkage, knowledge of various relevant data sources

Over to you: Causal mechanisms
Are you using an experimental or quasi-experimental design or appropriate qualitative approach to ensure the rigour of your results? If you are using 'internal' controls (i.e. recruited as part of your project) have you identified how to ensure an appropriate case:control ratio? If you are using 'external' controls (i.e. people who have not been identified through your activities) have you ensured appropriate access to reliable outcomes data? Can you identify what the effect size is and statistical significant of the effect? If you are using comparison or control groups, does your selection method take account of possible selection bias? Have potential contamination effects been addressed?

4.4.1   Strengthening a Type 3 evaluation

Aspects of good practice for a Type 3 evaluation encompass those noted above plus additional considerations. The following table summarises what is weaker and what is stronger evidence as part of a Type 3 evaluation, including best evidence.

The consideration of what counts as the best evidence advocates for more attention to be paid to the programme design. This may be especially important when you are considering using experimental or quasi-experimental methods. Attention also needs to be paid to the theory driving the evaluation. For example when your intervention effects are likely to be heterogeneous (i.e. varying across different cohorts or groups), looking at the overall effects can be misleading. In this case a more sophisticated design might be needed to identify when an intervention works for some and not for others.

Evidence is strengthened by using relevant outcomes, good research design, robust analytical strategies, recognising the significance and strength of any effects.

Dimension	Weak evidence	Developing Evidence	Example of best evidence
Outcome measures	Outcome measures not relevant to the aims of the activities.	Mechanisms for the collection of data that enables you to report the effect of your project on at least one key outcome measure which you can relate to what you are doing.	Mechanisms for the collection of data that enables you to report the effect of your project on at least one key outcome measure that directly encompasses notions of HE progression/take-up.
Research design	Research design not appropriate to the mechanisms/data available. Mixing of intervention and control group participants (cross-contamination).	Appropriate control group identified using rigorous quasi-experimental method. Control group members didn't participate in the activity, or otherwise benefit from it (i.e. no "contamination").	Randomisation of selection of treatment group and controls. Control group members didn't participate in the activity, or otherwise benefit from it (i.e. no "contamination"). Outcome measures undertaken in the same way, and at the same time between the treatment group and controls.
Project Design	No attempt to relate the outcomes measured to the processes involved in your project design.	Only a cursory attempt to explain the processes involved or to explain how the activities led to the observed results.	A well-developed understanding of the project design and the processes involved which enables it to be a replicable model with potential to inform future practices.
Analytical strategy	Analytical strategy inappropriate to the data.	Appropriate analysis competently undertaken.	Robust analysis of outcomes through the use of appropriate expertise.
Strength of the effect (effect size)	Outcomes are only changed very marginally (no more than 1%).	Moderate change in outcomes is observed.	Significant change is identified.
Demonstration of outcome	Tests of statistical significance of the difference between groups not undertaken.	Difference observed between groups is statistically significant at the 10% level.	Difference observed between groups is statistically significant at the 5% level.

Further help: Experimental and quasi-experimental methods

Resource	Comments	Available from:
Educational Endowment Foundation, DIY Evaluation Guide	A resource for teachers to introduce key principles of educational evaluation and provide guidance on how to conduct small-scale evaluations in schools involving pre and post tests.	https://educationendowmentfoundation.org.uk/tools/diy-guide/getting-started/

| Haynes, L., Service, O., Goldacre, B. and Torgerson, D. (2012) Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials. London: Cabinet Office - Behavioural Insights Team. | Part I of this paper sets out what an RCT is and why they are important. Part II of the paper outlines nine key steps that any RCT needs to have in place. | http://www.cabinetoffice.gov.uk/sites/default/files/resources/TLA-1906126.pdf |

Practice Example: RCTs of two information, advice and guidance (IAG) interventions delivered through text messages

Using an experimental research method such as an RCT is a very strong impact evaluation design. However, they tend to require specific circumstances, such as where it is possible to isolate the outreach activity to a particular group. Other challenges to be overcome are putting in place effective dedicated data collection from different groups, mechanisms for ensuring that appropriate background data can be obtained (with appropriate consent) and having sufficient sample sizes to complete a robust analysis. RCTs tend to work best for relatively simple interventions that are time constrained, and where they are conceptualised and set up as an 'experiment' rather than an ongoing intervention from the start. The following example looks at a large-scale IAG intervention using text messages. Issues of access to timely and relevant HE-access and progression-related information were identified as potential barriers to progression in education of high-attaining learners from disadvantaged groups.

Case study: The Network for East Anglian Collaborative Outreach (neaco)

Background: Two interventions were designed, one to guide Year 13 students through the UCAS application process by providing practical tips and guidance, including personal statement writing and finance; and one to offer Year 11 students information on post-16 options (and their implications for HE progression).[12]

How the evaluation was tackled: To allow for conclusions about causal effects of each of the interventions on students' application to HE (the main outcome measure for the Year 13 intervention) and self-reported knowledge of progression options (the main outcome measure for the Year 11 intervention), an RCT design was chosen. Recruitment and pre-intervention data collection for the two separate RCTs was undertaken as part of the main NCOP baseline survey, and a post-intervention survey is currently scheduled for autumn 2018. Post-intervention data collection for Year 13s was also carried out separately, by means of text messages, and a one-question survey. In total, 531 Year 13 students, and 810 Year 11 students were respectively included in the RCTs, with half of each year group randomly assigned to receive the intervention by gender (i.e. half of each group by gender and NCOP target status were randomly assigned to either the control or the intervention condition). Ethical considerations relating to withholding the text messages from the control group were addressed through the general level of school-based provision. A data linkage request has been lodged with the National Pupil Database to allow for further information (from consenting pupils) to be linked to their RCT data, and, eventually, tracked into the Higher Education Statistics Agency (HESA) record to obtain accurate information regarding progression to HE.

Results and learning: There was a relatively strong evaluation team in house due to a senior academic being seconded to the programme with experience in evaluation and statistical data analysis. External experts were consulted, including members of the Behavioural Insights Team to draw on previous experience of text-based RCTs[13]. Efforts went into ensuring high levels of school engagement in the data collection process. An event was hosted early to ensure schools' cooperation, where the RCTs were highlighted as important ways of obtaining good-quality evidence. Response to the main NCOP baseline survey (which included RCT recruitment) was incentivised at school level by means of a financial contribution for each respondent, sums which the schools could then use for activities to the benefit of the students. This enabled the evaluation to draw on a large-scale dataset (over 6,000 responses). Given the large scale of the project,

---

[12] Both interventions were delivered through between 15 and 20 text messages sent to consenting participants' phones. Where relevant the messages included links to further sources.

[13] https://www.behaviouralinsights.co.uk/education-and-skills/helping-everyone-reach-their-potential-new-education-results/

the evaluation activity currently amounts to under 5 per cent of total project budget; this includes the administration of the baseline survey, survey incentives, the cost of staff time, and also the implementation of the two tested interventions (which together make up just under 9 per cent of the evaluation cost).

How could the approach be developed? One weakness relates to the fact that these RCTs can only ascertain the impact of the interventions on the main outcome variables over and above any impact that may emerge from the regular school-based activity the RCT participants were engaged in as a matter of course through the NCOP. A further potential weakness of the approach is the possibility of 'contamination effects', i.e. control group participants benefiting from the intervention 'second-hand'. However, this is a tricky issue to address unless the intervention is confined to geographically distinct areas, which then potentially raises comparability issues.

5 Implementing your evaluation

Implementing an evaluation requires putting in place arrangements to manage and undertake evaluation, to collect or access appropriate data, to develop suitable data-collection tools, and put in place procedures and protocols to address ethical issues and ensure compliance with data regulations, as well as ensuring there are sufficient resources available to carry out the evaluation.

5.1 Planning for evaluation

An evaluation plan is a useful tool for planning the evaluation activities: the purpose is to set out the indicative evaluation activities that you intend to carry out in the different phases of the implementation. There are various possible approaches to developing an evaluation plan designed to provide an overall framework for ongoing evaluation and ensure that it is undertaken effectively and integrated as a management tool. It is up to providers themselves to decide when, how and what to evaluate, and this flexibility means there are likely to be various types of evaluation plans. The need for rigorous planning reinforces the importance of developing a culture of evaluation: evaluation not as an external exercise, but an instrument for learning and improving the implementation of your outreach programmes.

It is up to providers to decide what their evaluation plan looks like but be sure to have a well thought out plan in place.

Your evaluation plan should outline the evaluation activities to be undertaken, responsibilities for coordinating and undertaking and inputting to the evaluation, budget, any plans for oversight of the evaluation (steering groups, etc.), and arrangements for using results (dissemination, agreeing and monitoring recommendations). Once you have set out the steps that need to be taken to realise the evaluation you could then complete a risk assessment and put in place mitigation if needed. The evaluation plan should be an active document that is monitored and updated over the lifetime of the evaluation.

An evaluation plan aims to ensure the evaluation is undertaken effectively.

Over to you: Research strategy
Are the roles and responsibilities for managing, undertaking and inputting to impact evaluations clear? Have these been communicated effectively? Have any resource or capacity issues been addressed? Do you undertake risk analysis for your evaluations?

Practice example: RCT of online mentoring

Taking forward effective impact evaluation can be challenging, especially when it involves coordination across a wide group of stakeholders in a partnership. After you have decided on your evaluation plan you will also need to ensure that your partners are fully engaged and there are arrangements to ensure that the data can be accessed in an effective and consistent manner. It is often helpful to complete a risk assessment for the evaluation. A small-scale pilot could be used to identify the potential problems in advance so you can then put in place mitigation to address these.

Case Study: The Southern Universities Network (SUN)

Background: This project targeted Year 12 students in schools and further education colleges. It was designed to address the barriers to finding out about progression options and dispel the myths surrounding HE in geographically isolated areas through an eight-week online mentoring intervention.[14]

How the evaluation was tackled: The primary aims were to encourage Year 12 students in schools and their counterparts in further education colleges to progress to university, HE in further education (i.e. college-based Level 4 courses), Higher Apprenticeships and Degree Apprenticeships. Secondary aims of the intervention included raising HE knowledge, student aspirations and confidence/self-efficacy. In order to enable recruitment and randomisation for the trial, Year 12 students were invited to assemblies and provided with information about the intervention. Following expressions of interest, individual-level randomisation was carried out to allocate students to the treatment and control groups. Stratification was then carried out by school/college and sex. The research design involved measuring outcomes via a survey and follow-up at two points in time (spring and autumn).

Results and learning: A number of challenges were encountered during the implementation of the RCT that highlight best practice lessons for the future design and implementation of evaluations of outreach activities. The lessons point to the importance of:

- *Feasibility testing of evaluations that utilise experimental designs.* The budget and capacity implications of the planned RCT methodology could be tested through a small-scale pilot to pre-empt any potential problems that might need to be addressed before any large-scale trial is implemented.

- *Ensuring strategic buy-in from all parties involved in the trial.* Securing buy-in from the delivery partners is essential to the successful implementation of an RCT. This is particularly important for activities that engage with a range of providers including colleges across a wide area.

- *Setting realistic timelines.* It is important not to underestimate the time involved in getting Ethics Committee approval for the research, which can affect the data processes. It can also take considerable time and resources to disseminate initial information about the intervention, agree and disseminate survey questionnaires and instructions, and obtain consent and get documentation signed-off, which needs to be allowed for in the evaluation plan.

- *Establishing clear communication channels.* Ensuring that there is a single point of contact at the school/college is important to ensure that momentum is maintained for the trial and that accurate information is communicated on a timely basis throughout the evaluation.

- *Working with attrition due to drop-out.* Sustaining the activities was challenging for both the colleges and individual students involved. There may need to be over-recruitment of participants in order to ensure there are sufficient numbers of completers to run the trial. Having a named single point of contact in colleges and making sure there is regular communication between the partners can help to reduce the rate of drop-out.

How could the approach be further developed? This case study shows the importance of understanding the delivery context and processes involved in the provision of any outreach activity to understanding what works (as well as thinking about the outcomes for the participants involved). Further research is now underway looking at the processes involved in delivery, drawing on focus groups with the stakeholders and analysis of the patterns of drop-out from the project.

5.2 Data collection

Once you have decided on your outcome indicators at the programme-design stage you will need to consider what data or measures will help you to capture the changes your

Aim to specify what and how data will be collected at the

---

[14] The project ran from February to April 2018. The online mentoring platform provided tailored one-to-one advice and support about HE options, understanding the application process and personal statement development support.

intervention is seeking to make. Harries, Hodgson and Noble (2014) suggest that there are three questions you need to answer to create your evaluation framework:[15]

programme-design stage.

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

Tip: Before data collection begins decide what you need, why, how will you use it and what you will do with it during and after the evaluation.

The outcome indicators will determine what data is needed, which could include:

- Data collected specifically for the evaluation (e.g. surveys or interviews with participants and others such as teachers and parents). For example, you might collect information on the extent to which the intervention supported the outcomes you hoped for, e.g. did it affect future study or career expectations?

- Use of existing data. Evaluations of large-scale programmes tend to also use data that already exists or is being collected for other purposes for instance attainment data (performance in exams), progression data (post-16 destination data, applications and acceptances to HE, and student-record data.

New data will usually need to be collected for the evaluation as well as using existing data sources.

Sometimes you will need to collect data on 'proxy' measures if there is no direct measure of the outcomes you have identified. A proxy measure is one that stands when a direct measure is unavailable. For example, Free School Meal (FSM) eligibility is often used as a proxy for disadvantage. Often the indicators chosen to measure the outcomes of outreach will need to be translated into proxies in order to facilitate collecting data on them. For example, if you a seeking to find out about the value individuals attach to a university education then using questions such as whether they agree that the best jobs available in the labour market go to those who have been to university might be a proxy measure.

Proxy measures may need to be developed if there is no direct way of measuring the concepts you are seeking to capture.

Often new data-collection tools will be needed that can be used to collect data that is intervention specific (e.g. from your participants and stakeholders). It is critically important to be purposeful about designing your data-collection tools:

What do I want to find out? For example: questions could include
- How students' views of university changed during an intervention;
- Why students choose local universities rather than more selective universities further away

What data will best answer this question? For example:
- *Students' changing views of university:* The key here is that you are looking at change over an intervention period so a single data collection is not likely to answer your question. This could be investigated by surveys or interviews before and after the intervention; an audio diary with prompts maintained during the intervention; or a focus group.
- *Students' university choices:* This could be investigated through a task with a range of different types of universities, both local and national, for students to rank in order of preference and then to discuss choices; or through interviews; or an open-ended questionnaire.

The data collection methods and tools you put in place need to be appropriate for the outcomes you are seeking to measure.
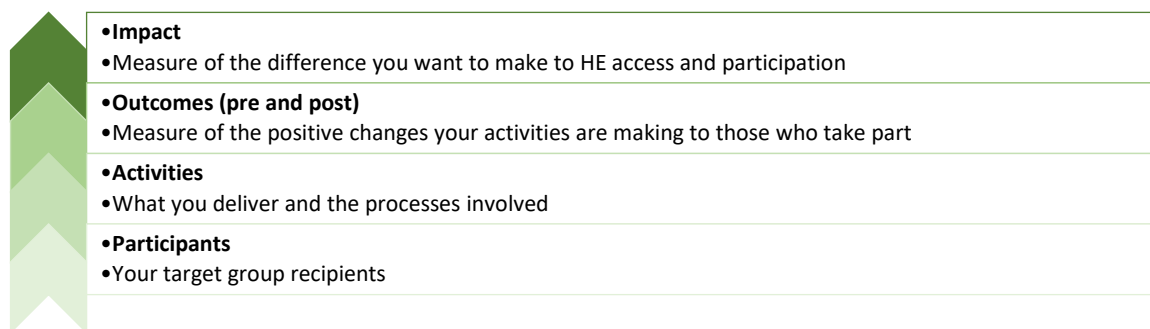
Well-rounded evaluation is informed by monitoring data that helps you to judge success given who was involved and how, and the context in which the outreach took place:

Monitoring helps you to assess who took part, in

---

[15] Harries, E., Hodgson, L. and Noble, J. (2014) Creating your theory of change: NPC's practical guide, London: New Philanthropy Capital, page 24. Available at: https://www.thinknpc.org/resource-hub/creating-your-theory-of-change-npcs-practical-guide/

- Participant data such as numbers taking part and background characteristics including targeting criteria data on any potential predictors of an outcome (such as socio-demographic factors and measures of attainment and motivation for HE).

- Delivery data such as information collected and used as part of delivery, describing the inputs and outputs (e.g. sessions provided and completed). Important details such as the type of activity, aims and objectives, duration and mode of delivery can help in an evaluation of what works best to achieve good results.

what way and in what context. It is important to assess the conditions in which outcomes were achieved.

- **Impact**
- Measure of the difference you want to make to HE access and participation
- **Outcomes (pre and post)**
- Measure of the positive changes your activities are making to those who take part
- **Activities**
- What you deliver and the processes involved
- **Participants**
- Your target group recipients

You should aim for individual-pupil-level data rather than aggregate measures, such as considering whole cohorts, unless there are a large number of cases. Individual data allows you to track how a participant in your activity is doing over time, for example, at the beginning of your project, just after your project and again a couple of years later. This allows you to infer whether your activity is likely to have made a difference to this individual. Individual data is also important when tracking across the student lifecycle.

Individual-level data is more useful than data for whole cohorts.

Tip: Consider using identifiers/names to link data and look at individual change. This will enable you to determine effects more rigorously and see what proportion has changed.

Response rate refers to the proportion of respondents participating in a data collection instrument such as a survey questionnaire. In longitudinal studies, it is usual for respondents to drop out of the study for various reasons (termed as 'attrition'). When using repeated data collection it is important to consider how many respondents you are aiming for and work out how many to start with to ensure sample size is not compromised by attrition. If you want to make generalisations, for example to say that results for a sample will be the same as for a wider population in general, then you need to make sure the data is based on a large group of participants.

The size of the sample will affect the inferences you can make from the results.

Skills required: Understanding of outcome and impact measures. Understanding of data-collection tools and techniques (e.g. questionnaire design). Familiarity with key widening-participation datasets. Understanding of data sharing and data protection rules and regulations.

Communication, partnership working and project management skills will also be important to ensure that the data collection processes are embedded and taken forward appropriately as part of the project design.

Many evaluations involve working with schools and colleges, and you might also need to make sure that agreement for data collection and data sharing is in place. If the evaluation involves working with schools, partnership agreements can help to ensure continued commitment from schools and to ensure data sharing is in place.

Partnership agreements can help to ensure access to data.

There could be challenges around the willingness of schools to adopt the framework of an evaluation project. For example, there may be reluctance from schools and colleges to engage where there is no guaranteed sustainability of the intervention. Similarly, while a randomised selection of students is one of the strongest designs, selecting a control cohort may prove difficult or present unacceptable ethical issues. Some providers have got schools to engage in evaluation by putting funding arrangements in place to pay for teacher time

Getting buy-in from schools can be a key challenge.

on data collection and evaluation (often alongside arrangement to fund delivery of outreach activities in school).

Tip: You may consider paying for the time of a teacher/teaching assistant; ideally this will be a strong champion and the right data person. Some providers use a partnership agreement with the school.

Further Help: Working with schools

Resource	Comments	Available at:
University of Exeter, Empowering Partnerships: Enabling Engagement Project	Regarding working with schools from a HEFCE Catalyst Project	http://socialsciences.exeter.ac.uk/education/research/projects/epee/
University of Exeter, Guidance for university staff	For specific guidance on working with schools. Provides an exemplar Memorandum of Understanding	http://socialsciences.exeter.ac.uk/education/research/projects/epee/researcherresources/understandingschoolcontexts/

Practice example: Working with staff in schools to embed data collection

Outreach where the delivery model involves staff based in schools and colleges can provide an opportunity to use school and college data to inform the impact evaluation. This could include gaining access to data for targeting (e.g. FSM status, attainment profiles) and also data that can be a proxy indicator for changes over time (e.g. improvements in grade predictions and attainment in exams where these are relevant to the type of outreach being delivered). Appropriate permissions to share and process the data need to be in place.

Case Study: The Access Project (TAP)

Background: TAP delivers an intensive, long-term intervention aiming at supporting young people from disadvantaged backgrounds access the top third most selective universities.[16]

How tackled: TAP's delivery staff are based in school, and identify students using schools' pupil attainment and demographic data, as well as qualitative input from staff in school (e.g. heads of year). The project team input and track student data manually, using a customised Salesforce platform to record individual level information (including participant monitoring data, assessments of student achievement (monthly); and end of Y11 GCSE attainment (annual)). This data is then analysed by an internal impact team, which conducts internal monitoring of programme outcomes throughout the year, and prepares the end of year evaluations.

Results and learning: Most of the work on evaluation is routinised (and has now run over several cycles). Project staff members in schools are able to secure data to inform analysis of academic attainment (part of their role is to input data on the latest assessments). A lot of information is collected on participants over four years. The central impact team stewards the IT systems, which collect data on an ongoing basis for case and outcome management (as well as evaluation). The IT systems are developed internally to ensure they are fit-for-purpose. It helps that the evaluation team are former access officers who understand the constraints and opportunities for data collection plus the data supports day-to-day operations so the project officers understand its importance. The data is capable of teasing out differences, e.g. differences in average scores between schools, which can be a starting point for unpacking contextual and delivery issues. This has been possible because TAP has been working since 2008 on a relatively stable programme, and has developed knowledge on what's important to capture (and ways of quantifying it) based on their learning over time. The evaluation evidence has informed decisions about the intensity and timing of the delivery.

How could this be developed further? There is scope to look at the outcomes of the work in a nuanced way, taking account of the context and personnel involved in the delivery. This approach would be important in helping consideration of issues of consistency in the quality of outreach delivery.

Over to you: Data collection

---

[16] The programme is set up to tackle two main barriers to access: academic attainment (at GCSE and A-level) and university readiness (defined as the set of knowledge, attitudes and skills that ensure the successful navigation of the HE admission process).

5.3　Ethical evaluation and Data Protection

There needs to be a lawful basis for collecting and processing personal data (i.e. information about a particular identifiable individual). The General Data Protection Regulations (GDPR) set out key principles and obligations for processing personal data.[17] You must be able to demonstrate that processing of personal data is necessary for a specific purpose. Important considerations include:

Data protection regulations need to be applied.

Obtaining consent

It is extremely important to obtain participants' consent for involvement in the evaluation and the data that will be collected, not only to provide a legal basis for processing personal data but also for ethical evaluation. You must ask people to actively opt in, and should make clear your organisation and any third parties who will rely on the consent, why you want the information, what you will do with it, and that individuals can withdraw their consent at any time. This could be in the form of an information sheet accompanied by an opt-in form for prospective participants. Ideally this should include what data will be collected and for how long it will be stored. It is helpful to let respondents know if a follow-up data collection exercise is planned.

Opt-in is preferable to opt-out.

Anonymity and confidentiality

The participants should not be identifiable in any reports and publications. This is important as it can often leave people in a vulnerable position. For example, if there is a possibility that individuals could be identified, the utmost caution needs to be taken to maintain the anonymity and confidentiality of data held by the organisation (usually quantitative data relating to less than five individuals is suppressed and any case studies are anonymised).

Data should be confidential and steps taken to anonymise data in reports.

Data linkage

The data for an evaluation can come from different sources and it may be necessary to merge datasets. Merging datasets requires care to check mismatches which can ruin the analysis and hence the evaluation. In the case of official datasets, linkage is taken care of by Government departments and the evaluator receives an anonymised merged file with reduced sensitivity as agreed by the data-approving panel.

Data linkage presents particular challenges.

Transparency

All individuals and agencies involved should be fully aware of the nature of the intervention, planned data collection and use of data from the start (including data sharing where relevant). This kind of transparency ensures the interests of all parties involved are taken care of and privacy requirements are met. Participants can ask for their data not to be included in the evaluation, or withdrawn at any time they wish. Also, they should be aware that they can ask for a copy of any evaluation findings.

There should be complete transparency on data use and participants have the right to withhold/withdraw participation.

Storage and security

Major concerns are sensitivity of the data and security measures. Only required information should be collected. It is possible to reduce the sensitivity of data if the individual cannot be identified by the information available. For example, if date of birth and six-digit postcode are not required for the evaluation they should not be collected. In general, sensitive personal data should be encrypted and should not be shared.

The amount of sensitive personal data should be minimised.

---

[17]For guidance on the requirements of the GDPR see: https://ico.org.uk/for-organisations/guide-to-data-protection/introduction-to-data-protection/introduction-to-data-protection/

It is important to consider how data is stored, where, and who is responsible for it. A data management plan that meets data security and privacy requirements should be used. You must consider how long data is kept and the process for destroying it. [18]

You will need a plan for data storage, security and data disposal.

Other ethical responsibilities to participants

Before collecting data, researchers have a responsibility to think through the duty of care in order to recognise and prepare for any potential risks, and should seek to minimise any harm that may arise. In the case of research with young children and those in vulnerable circumstances, researchers should get the approval of those responsible for such participants. If you use incentives to encourage participation, the level of incentive should not impact on the free decision to participate.

You should aim to manage any potential risks.

Further help: Ethical guidelines and data protection

Resource	Comments	Available at:
British Educational Research Association [BERA] (2018) Ethical Guidelines for Educational Research, fourth edition, London: BERA	An authoritative summary of ethical issues associated with undertaking an evaluation research project in educational contexts.	https://www.bera.ac.uk/researchers-resources/publications/ethical-guidelines-for-educational-research-2018
Information Commissioner's Office	Information on data sharing, tools and resources	www.ico.org.uk

5.4 Validating data collection tools

The methods and tools you use to collect outcomes data might vary but to be valid need to be suitable to answer your impact evaluation questions (i.e. looking at exam results, for example, is helpful to measure progress in education but might be less useful if you want to know about attitudes to HE). You may need to develop tailored methods for collecting data, for example through questionnaires. When developing evaluation tools such as questionnaire surveys, you should test whether the responses are valid in terms of what you are trying to find out. It is good practice to pilot the survey in advance, and it is also helpful to use qualitative research, for example a focus group, to test respondents' understanding of the questions you are asking.

The data collection tools need to be suitable to answer the question(s) you are seeking to answer.

Key Term: Validity

Validity refers to whether a data collection tool (test results, surveys, questionnaires, logs etc) measures what it is supposed to measure accurately.

It is particularly important to think about the validity of the data collection tool when you are using 'proxy' indicators. For example, using improvements in attainment as a proxy for success in HE would only be a valid measure if you can demonstrate a strong relationship between the two.

The process of validating a data collection tool generally involves cognitive testing in the first instance. Statistical tests and measures can be used to assess whether data collection tools that generate quantitative data, such as surveys, are generalisable.

---

[18] The Information Commissioner's Office highlights the role of a data protection officer and steps to be taken if a data breach occurs. For a more detailed discussion please see https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/data-protection-officers/

Depending on the aims and context for outreach, there are some pre-existing validated tools that might be relevant to some types of interventions (for example, measures of well-being). However, it is likely that these may be too general to be usefully applied to widening-participation outreach which tends to be highly contextualised and HE-focused. If you are designing your own surveys or other types of data collection you need to make sure they are administered reliability to ensure consistency in the interpretation of the results. When you develop your data collection plan, you have to create something that is do-able in your context. You may have to make decisions about what you can afford to do in terms of data collection and analysis. If you do have to make compromises, it is better to go for depth, rather than breadth, i.e. try to gather high-quality data about fewer outcomes, rather than more superficial data about all the outcomes.

Tools such as surveys need to be administered reliably and consistently, and should be piloted to make sure the questions are interpreted correctly.

Tip: If you are interested in changes in attitudes, consider tying your evaluation to well-established psychological or sociological constructs such as self-efficacy or social capital. Use pre-validated tools, or ensure you cognitively test your own.

Further help: Cognitive testing questionnaires

Resource	Comments	Available at:
Lenzner, T., Neuert, C. and Otto, W. (2016) Cognitive Pretesting, GESIS Survey Guidelines, Mannheim: GESIS Leibniz Institute for the Social Sciences	Provides an introduction to questionnaire pre-testing, focusing on methods and techniques you can use.	www.gesis.org/fileadmin/upload/SDMwiki/LenznerNeuertOtto_Cognitive_Pretesting.pdf

Practice example: Quantifying personal change

Many widening-participation researchers have drawn on Bourdieu's concepts of intellectual capital (including subject expertise), academic capital (understanding of tacit 'rules of the academy') and social capital (social connections). Measuring attitudes, aspirations and other psychological constructs such as self-efficacy can often be slightly tricky and may call for focused research. It may involve quantifying data relating to qualitative concepts or changes.

Case Study: The Access Project (TAP)

How was the evaluation tackled: TAP has articulated, at a student-level, the long-term, intermediate and short-term outcomes the programme seeks to achieve and has developed a range of monitoring and evaluation tools to allow for both robust impact measurement and effective impact management. The measures reflect the main delivery objectives i.e. improvements in: academic attainment (at GCSE and A-level) and university readiness (defined as the set of knowledge, attitudes and skills that ensure the successful navigation of the university admission processes). Information is collected 'live' during sessions with students and through surveys at three points in the programme (baseline, mid-term and final year). TAP has introduced a new element this year, specifically assessing young people's self-efficacy. A set of tools for measuring student self-efficacy at key points on the programme is being tested – TAP is using this as an intermediate outcome to assess students' progression in terms of university readiness.

Results and learning: Previous analysis of the data showed that the type of readiness indicators correlating to good outcomes were measures of independence, and this has put the focus on further work to develop individuals' independence scores during one-to-one sessions. The increasing focus on self-efficacy has developed alongside TAP's articulated Theory of Change, and learning from the interventions which scaffold progression outcomes at a detailed level of granularity (the analysis encompasses over 70 variables and uses small subgroup analysis). It is expensive to resource to maintain this level of overview and TAP is fortunate to have the backing of funders that support the organisation to upskill their impact ability through core funding. Rather than only working to outcomes that are easy to measure, the approach is working to go beyond summative evaluation by looking at the underlying elements of success. Talking and listening to the practitioners helps to ensure grounding in reality while avoiding over-simplifying a complex picture. The key

includes identifying a baseline and triangulation of the evidence. Self-reported measures tend to have a poor reputation, but their usefulness can be increased when viewed in combination with other evidence. By putting in place a well-thought-out process to develop and validate measures of self-efficacy TAP Is hoping to build an evaluation tool capable of teasing out impacts over time.

How could this work be further developed? This endeavour has required developmental work and a long time scale: building the case; reviewing available tools; developing a new tool specific to the HE context; collecting pre and post intervention data to validate the tool; and testing the results with a pilot group (forthcoming). Having access to a large student population (800-1000 replies in Years 11, 12 and 13) enables detailed statistical analysis to underpin the validation process.

5.5 Tracking participants

It is useful to understand the final result the activity is aiming for. Is it about improving educational outcomes or is it about raising expectations to progress in education? Educational outcomes for example can be measured by performances in standardised national tests such as at the end of Key Stage 4 or 5 examinations or even by tracking post-16 participation in certain identified subject areas. It then becomes crucial to have some element of longitudinal tracking to assess the impact of the programme.

Tracking over time is usually required in order to establish educational destinations and achievements.

It may not always be possible or necessary to collect primary data directly from participants, if this is available through existing administrative datasets. For example, school and pupil-level information could be available from administrative datasets in schools, or via the National Pupil Database (NPD), although this data often takes time and specialist expertise to access. UCAS and HESA data provide information on HE applicants and entrants. This kind of secondary data can prove to be a very helpful data source for outreach activity providers as these are populated via census or hosts and this information is regularly updated. Plus, using pupil-background indicators can significantly improve the robustness of the evaluation design.

Secondary sources of data can be useful, although there can be issues in terms of negotiating permissions to accessing the data.

Different approaches to accessing data on outcomes over time are summarised below. Depending on the source, you will need to make sure that appropriate data-sharing agreements and permissions to use linked data are in place.

Data source	Benefits	Disadvantages	More information
Implementing follow-up surveys with participants (telephone or postal follow-up)	Allows you to find out about outcomes in a direct and timely way. You have the possibility of collating views/ perceptions of the influence of the intervention as well as data on the outcomes achieved.	Following up participants can be resource intensive, depending on the numbers involved and methods used. Response rates can be low and there is a problem of response bias.	
Using your own datasets (for example on applications to undergraduate courses)	Available 'in-house' Quicker to obtain (sometimes in 'real time')	May miss outcomes for participants who progress to other HE providers	Consult with your data services/management information team
Using partners' datasets (e.g. via schools or local authority sources)	Local sources are usually more direct and you may be able to negotiate on the types of information which is fit-for-purpose for your needs	Requires data sharing processes to be in place Data from different sources (e.g. multiple schools) can be difficult to aggregate	Information on data sharing at www.ico.org.uk
Linked administrative data – data extracts from the NPD, Individualised Learner Record (ILR) and HESA student record.	Very comprehensive source which includes contextual factors as well as outcomes	You may need to put new processes in place for securing and working with the data. There is a time delay in receiving data (typically six to	DfE manages the application process, see: www.gov.uk/government/ collections/national-pupil-database

		12 months after the data is processed)	
Tracking systems (HEAT, EMWPREP, Aimhigher West Midlands)	Shared service facilitating access to regular longitudinal datasets	There are costs involved. Time lags mean data will not be immediately available. Once you have the data you may still need skills to analyse and interpret it	www.heat.ac.uk www.emwprep.ac.uk www.aimhigherwm.ac.uk
UCAS STROBE	Access to applicant tracking data and comparison group data at aggregate level. Application data is available earlier in the student journey than via the student record.	There are costs involved Based on applications and acceptances rather than student progression outcomes	www.ucas.com/data-and-analysis/data-products-and-services/strobe

Tip: Try to avoid over-reliance on tracking as the main focus of the evaluation. Tracking can provide important data to inform the evaluation, but well-developed evaluations would aim to provide multiple perspectives on the influence of the activity on the outcomes.

Over to you: Data collection
Do your participant data collection arrangements allow for measurement of individualised change (as well as cohort or subgroup analyses)? Have you established a methodology to track the outcomes of your participants over time? Do you obtain data using validated or sector-standard tools and techniques?

Practice Example: Systems for longitudinal tracking

For schools outreach work it is often a challenge to put in place systematic data collection that sustains across years, particularly where the outreach is with younger students for whom it is going to be hard to establish any direct connection between outreach and education progression in later years. At the same time it is still important to try to get evidence on whether this kind of outreach is having any effect in order to inform future activities.

Case study: Loughborough University

How tackled: The University of Loughborough hosts and is part of a widening participation monitoring and evaluation partnership, EMWPREP (East Midlands Widening Participation Research and Evaluation Partnership). Other providers are collaborating together to develop systems for storing their widening participation monitoring data and linking to administrative datasets (including HE outcomes) via HEAT.

Results and learning: The EMWREP system for example allows recording individual participants' interactions with Loughborough and other EMWPREP partner HEIs (subject to consent from parents). This system also allows monitoring of key information such as an individual's background characteristics (gender, ethnicity, first in their family to HE, disability, care status). This data can then be used to analyse differences in attainment and progression to HE.

How could the approach be developed further? There is potential for developing pair matching/matched cohort methods using the large-scale administrative datasets (for example, using PSM techniques). This work would require a data extract from the NPD (linked to student outcomes) and results available at aggregate level to allow an assessment of the likely impact of the outreach on the participants involved.

5.6   Resources for evaluation

Different types of resources are usually used during an evaluation, and the evaluation workplan should set out the resources required which could include:

▪ Financial, e.g. costs involved in collecting data such as undertaking interviews;

Resources for evaluation include not only the direct

- Management resources, e.g. day-to-day oversight of the evaluation tasks and quality assurance;

- Technical expertise, e.g. analytical input into the evaluation design, expertise for data processing and data analysis;

- Partnership working, e.g. liaison with other organisations and individuals involved in delivering the intervention and/or collecting data in order to consult on the evaluation and communicate on the implementation and findings;

- Stakeholder inputs, e.g. securing involvement of other stakeholders (such as people and organisations directly affected by the intervention;

- Peer oversight, e.g. securing external inputs to the research design and methodology etc., for instance by setting up an evaluation steering group;

- Other costs such as subscriptions, costs of specialist software for data analysis etc.

costs and technical expertise but also the time involved in managing relationships to ensure the evaluation is delivered appropriately.

The indicative costs of impact evaluation are usually considered in proportion to the delivery costs. Typically, for interventions that are innovative (especially pilots for new delivery methods) and where evaluation is needed to inform learning, the costs are likely to constitute a larger proportion of programme resources (at least 5 per cent and possibly more where evaluations are part of the management's implementation strategy, for example, where evaluation includes a strong formative element). It is useful to indicate at least a minimum budget for direct evaluation costs. However, the budget is likely to depend on the tasks that are included and the methods that are used to perform the evaluation – i.e. the nature and scale of the evaluation. The resources required may also depend on the nature of the outreach intervention being evaluated. As a rule of thumb the least intensive and most well proven interventions will require less evaluation resource than more intensive and innovative interventions.

There are no hard and fast rules as to the level of resources required for evaluation, as this depends on the complexity of the task.

Senior managers and decision-makers are well placed to ensure that resources for evaluation are prioritised. It is helpful if institutional resources are deployed with evaluation in mind (for example, by including budget lines for evaluation as a matter of course in project delivery plans and establishing protocols for the indicated level of resources for impact evaluation as a share of project budgets).

Tip: The cost of evaluation is countable in your A&P Plan

Ideally a budget will have been secured for evaluation as part of the planning stage, proportionate to the type of activity and needs of the evaluation. Doing preparatory work within the development phase of the programme, e.g. building in mechanisms for information collection as part of the delivery) may help to reduce the evaluation costs. It is important not to let resource constraints be a limiting factor on the quality of evaluation. If there is limited resource then you could consider the following:

You should aim to set a budget for evaluation and secure the human and technical resources required.

- Having a tightly focused evaluation to minimise the cost. You should focus on capturing the measures of success which have the highest priority in relation to your strategic objectives for the outreach;

- Using a sampling method to cut down on the amount of data required from across a very large number of participants;

- Utilising cost-effective approaches (for example, running beneficiary surveys online or having telephone rather than face-to-face interviews, reducing the amount of questions or consultations, building data collection into your delivery);

- Asking academic staff in your organisation to offer their expertise to support the evaluation, or draw on Masters and PhD students as researchers.

Costs can be minimised if necessary by having a focused evaluation, building data collection into delivery and utilising internal sources of expertise.

Suggestions for practice:

The indicative costs of evaluation are usually considered in proportion to the delivery costs. Typically, for interventions that are innovative (especially in the case of pilots for new delivery methods) and where evaluation is needed to inform learning, the costs are likely to constitute a relatively high proportion of programme resources (at least 5 per cent and possibly more where evaluations are effectively part of the management's implementation strategy, for example, where evaluation includes a strong formative element). In general intensive and innovative interventions will require most evaluation resource.

It is probably useful to indicate at least a minimum budget for direct evaluation costs. However, the cost is likely to depend on the tasks and methods used to perform the evaluation – i.e. the nature and scale of the evaluation, and therefore resources required, will depend on the evaluation design and intervention being evaluated. It is especially important to ensure time and resources are available for evaluation data collection and liaison on the evaluation requirements where there are multiple partners involved.

6   Learning from evaluation

6.1   Use of evaluation

When planning an evaluation it is useful to give some thought to how you will use the findings. Ideally the evaluation will be part of the project planning cycle. Either way getting the evaluation results is not the end of the story: rather the evaluation report should generate conclusions that should be discussed and transformed into actions by the outreach provider and partner stakeholders.

Impact evaluation is part of the project planning cycle.

The Project Planning Cycle
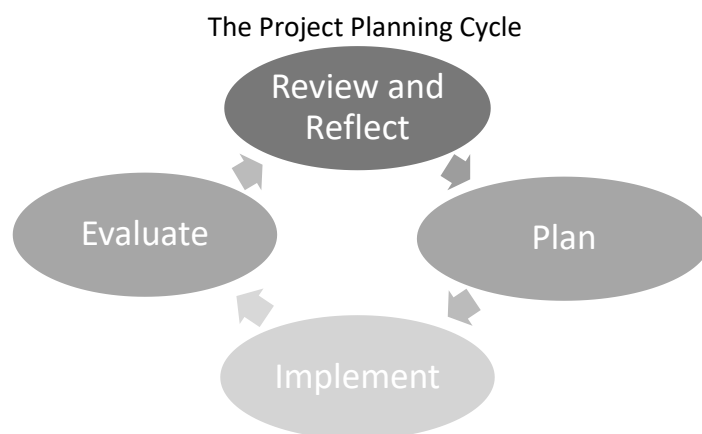
Tip: Evaluation is an integral part of outreach programme planning and management. It is not purely an external imposition. Managers should think of evaluation as a resource: a tool for improving your success and a way of systematising knowledge about effective outreach practices.

Suggestions for practice

Useful mechanisms to ensure that the organisation benefits from the learning generated by the evaluation include:

- A strategic plan for widening participation that specifies how your evaluation programme helps you to understand whether you will meet your aims and objectives.
- Clarity on the arrangements for using evaluation (for example, setting out in your evaluation plan how you will disseminate the results, and agree and monitor any recommendations emerging).
- A commitment to continual improvement of the effectiveness of an activity in its context through an ongoing cycle of review, consideration and revision.

6.2 Interpreting results

All types of evaluation can contribute to learning about the effectiveness of outreach and can highlight any aspects that are less effective or require reconsideration of an intervention. If the results are available as part of the outreach activity planning cycle, the improvements can then be made in the next round of delivery. The different types of evaluation provide different types of evidence, and this affects the level of confidence in the results in terms of knowing whether or not the outcomes and impact were the result of your outreach activities. This will also affect how you present and use evaluation when making claims about the effectiveness (or otherwise) of what you are doing.

Different types of evaluation provide different evidence from which to draw conclusions.

Type 1 Evaluation: Narrative	Type 2 Evaluation: Empirical research	Type 3 Evaluation: Causality
We have a coherent explanation of what we do and why Our claims are research-based	We can demonstrate that our interventions are associated with promising results	We believe our intervention causes improvement and can demonstrate the difference against a control or comparison group using an appropriate research design

As well as analysing data to describe what happened, Type 3 and many Type 2 evaluations usually investigate questions or test hypotheses e.g. did the participants in the outreach have better progression outcomes than they would otherwise have done without the intervention? Inferential statistics make a judgment about whether or not any observed difference is probably down to the intervention. Of course the claims you can make will also depend on what the evaluation work finds particularly in relation to the observed changes for your participants compared to a comparison or control group. For example, when analysing quantitative results when you have undertaken multivariate analysis you may have three possible findings from your statistical analysis. These are summarised below.

The conclusions you make should be informed by the results.

Result	Possible conclusion
Result A: The targeted group of participants have significantly **better outcomes** than the comparison group.	This suggests that your activity is effective at improving outcomes or you are seeing the effect of an unobserved bias in your intervention sample.
Result B: The targeted recipients have the **same outcomes** as the comparison group.	This can suggest that there is actually no impact of your activity on the outcomes you measure or there is a contamination in the control group leading to a *dimmed effect.*
Result C: The targeted recipients have significantly **worse outcomes** than the comparison group.	This indicates that your activity is detrimental for the target group.

In drawing conclusions from your data analysis it is important to apply the following tests:

Is the conclusion reasonable?

For instance, if you're doing an evaluation of whether outreach participants achieve better results in exams following participation in a master class intervention (i.e. that there is a positive relationship between a master class and exam results) you would want to show that the intervention gave enough input to make this relationship credible.

Is the conclusion valid?

You might be able to show that participants get higher exam results than those who didn't take part, but we might conclude the intervention was not the cause. Some other factor could have caused the difference (e.g. differing levels of prior attainment). In that case the conclusion that the master class had an effect is not valid because there is an alternative explanation for the observed result. In order to demonstrate your intervention caused an outcome you need to 'rule out' other possible causes, which is why a good research design aims to take account of other possible explanations.

Does your study measure what it intended to measure?

When you implement an activity you will have a theory about how it will work. A student ambassador programme may be designed to provide access to information about HE. However, it could be that participants benefit from individual attention from the ambassador. Your evaluation should aim to demonstrate the relationship between the activity and the outcome in order to test your intervention theory that access to information has made the difference to the participant.

Can the conclusion be generalised?

Impact evaluation is usually completed for a certain intervention, with a particular target group, in a specific delivery context, at a point in time. It is therefore useful to think about the extent to which the participants involved, the delivery context (e.g. set of relationships) and the timing affected the results that were achieved. Unless you are able to demonstrate that the intervention got the same results with lots of different people in different places at different times, these issues should be explored in your evaluation report in order to specify the conditions under which your intervention was successful.

Aspects of quality should be applied to how you report the results of your impact evaluation. The aims should be to ensure objectivity, accuracy of reporting of the findings, and transparency on what the evaluation involved and the resulting robustness of the evidence.

Evaluation as a discipline requires objectivity i.e. efforts to ensure that the evaluation is objective and there is appropriate challenge and scrutiny of the activity. There is perhaps an inherent tension in providers undertaking a self-evaluation – because practitioners usually want to show their work in the best light. Sometimes objectivity requires working with others in your organisation or external evaluators to ensure the evaluation is as objective as possible (as well as providing access to expertise). You could also set up a steering group of external experts and stakeholders who can play a role in bringing scrutiny. For example, a project officer might not be the best person to evaluate if they are immersed in the project and less objective than someone who is less heavily involved with the programme. It is important to consider if your approach avoided building in bias. Applying standards to what you do should help to promote objectivity.

You should be transparent about any weaknesses in your evaluation methodology, research design or data collection in your evaluation report. Ideally your evaluation report should contain sufficient technical detail for others to judge for themselves the robustness of the findings. As noted in Section 4.3.1 you can enhance the chances of statistical analysis providing powerful results by boosting the number of cases, ensuring the data is good

You should ask yourself whether the conclusions you are making are reasonable considering the quality of the evidence you have generated; that your assertions of impact are valid; that your measures are appropriate to what you are trying to achieve and whether the conclusions depend on any particular set of conditions.

Evaluation reports should be objective, accurate, and transparent.

You have a choice of who should do the evaluation, and should ensure they are unbiased and objective.

Make sure you recognise any limitations in the

quality (i.e. as comprehensive as possible with few missing variables), and using as reliable as possible measurements (i.e. validated tools or objective measures).

evaluation design or implementation.

Evaluation activity is about both proving and improving, which means highlighting where impact was not achieved yet or the ambitions were unrealistic. Evaluation helps providers and the sectors to avoid investing in activities that do not create the outcomes and impact for the target groups that are desirable to improve access to HE. In this respect evaluation is a source of lessons learned, especially when evaluations take place during an intervention so that changes can be made before it is too late.

Negative results are just as important for learning as positive evaluations.

Over to you: Interpreting findings

Does your evaluation reporting acknowledge the limitations of the research design approach used in each case? Can you attribute the impact – or lack thereof – to the intervention? Does your evaluation triangulate findings from different sources? Does your reporting demonstrate engagement with the scholarly literature on effectiveness where it exists?

Suggestions for practice:

Take a critical approach to thinking about the evaluation you are undertaking, for example, by recognising any issues relating to the method, sampling, or other issues such as selection bias.

Where you have evidence that an intervention is associated with a positive change in outcomes for your participants, make sure that it is reasonable to expect your activity to have contributed to this, for example, by gaining a clear sense of the factors and processes involved in bringing about the change.

Avoid an over-reliance on one source of evidence, especially participant self-reported data. Ideally your evaluation should draw together data from different perspectives, for example, from teachers or other professionals as well as students.

Aim to show how your results contribute to the body of knowledge about the impact of the activities you are undertaking, for example, by making reference to how your evaluation approaches have been informed by existing studies of impact and how your results compare.

Case Study: Testing the programme theory underpinning an outreach activity

Practice Example: Rolling out reflective practice at Go Higher West Yorkshire (GHWY)

Background: GHWY delivers tailored, place-based interventions, supported by programme theories, which articulate the assumptions underpinning the work. The programme theory for the community strand is based on the assumption that reaching into a young person's community will change culture and support a young person to take the step into HE. Changing culture within the communities, and with outreach providers which have traditionally put the focus on working in schools, is not something which happens overnight or which can be easily measured.

How the initiative developed: The GHWY evaluation framework applies a realist methodology to seek to understand 'what works' *for whom*, *in what circumstances* and *why*. The approach to evaluating the community strand sought to be sensitive and nuanced to the community and provider contexts for the outreach. By opening up a dialogue between the evaluation team and outreach staff delivering this work, and listening to some of the challenges it presents, the GHWY team have been able to learn from where things have not worked so well, as well as the successes. The collection of evidence has been strengthened by treating community outreach as a form of research, which required an effective means of capturing data. The evaluators created a reflective journal for outreach practitioners delivering in the community (based on reflective workbooks created for outreach participants), which provided a non-judgmental space in which to record those activities that have worked less well.

Results and learning: Enabling staff to approach this area of work with an experimental focus created a breakthrough in understanding and practice. By collating intelligence about what doesn't work as well as

what does, the approach informs the strategy and helps to target future activity where it might be more effective. Reflective practice has been normalised, which has encouraged delivery staff to take the time and space to think through *how*, *why* and *in what context,* an activity has worked or not worked. Reflective practice is now being applied to other activities and the reflective templates have been rolled out to teachers and school staff attending events with their students. Alongside participants' reflections, and evaluation data collected via pre and post surveys, GHWY now has valuable session-by-session feedback from key influencers. This is a low-cost evaluation initiative, which provides rich data from a range of stakeholders, and deepens understanding of the contextual factors that impact outreach delivery. It provides a crucial narrative on which to base delivery and an evaluation that aligns with the wider GHWY evaluation framework.

How could the approach be developed? Recognising the value from a research and evaluation perspective of insights from delivery staff and stakeholders such as teachers may seem obvious, however providing a structured format has helped to capture learning to test the assumptions underpinning the activities and to inform practice. GHWY has started to use practitioner focus groups and interviews with staff involved in community outreach, alongside other evidence, to continue to understand how delivery can be improved. Surprisingly, in light of the challenges at the outset, the community strand is proving to be an area of strength.

6.3   Sharing results

Another important feature of using standards is to support the communication of evidence, in order to ensure that evidence is not only produced, but that it is accessible, and rigorous enough to influence future practice. As well as the internal audiences for your evaluation it is important to consider how you will present the findings more widely. If you commission the evaluation externally you should specify what kind of report(s) and dissemination you require as part of the commission, and even if the evaluation is completed internally, you should still think about what the reporting arrangements will look like. Examples of evaluation outputs include: a full evaluation report, technical report, executive summary, briefing paper, academic article, conference or presentation materials. It may be that your institution has protocols and local arrangements for publishing information including the results of evaluations.

Think about how impact evaluation results can be shared in order to influence future practice internally and externally.

Tip: Having a clear idea of how you will use the results of the evaluations to inform your future outreach work will help you to think about what type of analysis you will need.

The findings from evaluations should also be fed back into the policy process as part of your reporting on you're A&P Plan. The results from evaluations, the recommendations emerging or what you will do as a result of the evaluation based on the learning, should inform future decision making on outreach.

Report evaluation results to OfS as part of the A&P Plan process.

Lodging evaluation reports in a repository, such as that being developed by the Evidence and Impact Exchange,[19] is important to allow those who might want to use the results to do so, and to enable future use of the evidence in a systematic review or meta-analysis.

The aim is to build up a repository of evidence of impact of outreach.

One of the biggest concerns, as well as availability of evidence, is the possibility of bias in which evidence is made available and which is not. In particular, there is a concern that practitioners are more likely to share evaluations that show their intervention in a positive light. Therefore there is potential for the evidence base to be distorted. Unless inconclusive

Inconclusive results and negative

---

[19]   Further details at: https://www.officeforstudents.org.uk/advice-and-guidance/promoting-equal-opportunities/using-evidence-and-evaluation-to-improve-outcomes/evidence-and-impact-exchange/

evaluation results and evidence of ineffective practices are shared, evidence users will not be able to make decisions based on all the evidence than has been produced.

evaluations are also useful.

Tip: Wherever possible evidence should be made available and accessible, whether the findings suggest the outreach is effective or not.

Sharing information on what does not work is as important for learning as positive evaluations. The OfS encourages provides to share information on their failures as well as successes because the most important thing is to ensure future investment in access and participation measures is effective as possible.

Another issue is making sure that the reporting includes enough information on the evaluation methodology to enable evidence users to understand the study and assess the credibility of the evidence. Your evaluation reports should identify any limitations with the evaluation design (common limitations tend to be in terms of the sample sizes involved or difficulties in controlling for selection bias). You may also want to consider publishing the data analysis alongside the report. This can help to enhance the credibility of the results and could potentially enable researchers in future to combine information from multiple studies to generate new findings (known as 'meta-analysis').

Be sure to provide enough information on how the evaluation was undertaken to enable others to use your evidence appropriately.

Over to you: Sharing results from evaluation

Do you have a mechanism in place to share the findings from your evaluation internally? Are mechanisms in place to enable evaluation results to influence practice across the sector? Is your evaluation contributing to the body of knowledge held by the Evidence and Impact Exchange?

Suggestions for practice:

Internal mechanisms for sharing the results of evaluations of impact of outreach could include cross-institutional networks or steering groups. Ideally a structure should be in place to oversee any actions agreed as a result of using evaluation evidence.

A wide range of opportunities can be used to share the results of evaluation externally in order to contribute to knowledge on effective outreach practices, such as publication of reports and briefings, presentations at conferences/events, publication in widening participation newsfeeds, and articles in journals. In order to maximise the influence of the evaluation a dissemination strategy could be developed. Examples of reporting formats and content suitable for different audiences might include:

Audience	Content	Format(s)
Policy makers	Quantifiable results that enable comparisons with other interventions and providers	Summary report, information included in monitoring reports
Widening participation practitioners	Information on the evaluation and lessons learnt	Workshop presentation, newsletter articles, blog post
Partner stakeholders	Evaluation results and emerging recommendations for developments or changes to the intervention	Detailed written report

7   Where can I get further support?

Access and participation standards of evidence

This document lays out access and participation standards of evidence for senior managers, decision-makers and practitioners with a remit for evaluation and reporting on access and participation activities. The aim is to promote understanding of the standards of evidence and a more rigorous approach to undertaking and using impact evaluation to improve the effectiveness of the investment in access and participation programmes.
See https://www.officeforstudents.org.uk/publications/standards-of-evidence-and-evaluating-impact-of-outreach/

An evaluation self-assessment tool

The self-assessment involves reflecting on your approach to evaluation against a series of questions. This tool has been developed to assist providers to review whether their evaluation plans and methodologies go far enough to generate high-quality evidence about the impact of their activities in the A&P Plans, highlight areas for potential improvement and facilitate benchmarking across providers.
See https://www.officeforstudents.org.uk/advice-and-guidance/promoting-equal-opportunities/evaluation-and-effective-practice/standards-of-evidence-and-evaluation-self-assessment-tool/evaluation-self-assessment-tool/

Crawford, C., Dytham, S. Naylor, R. (2017) The Evaluation of the Impact of Outreach: Proposed Standards of Evaluation Practice and Associated Guidance, Office for Fair Access

This document provides a summary of evaluation principles and key stages in the development of evaluation strategy, sets out the types of evaluation of the impact of outreach, and provides guidance and worked examples.

Box 1: Resources for people wanting a general introduction to evaluation		
Resource	**Comments**	**Available at:**
Evaluation Capacity Building (ECB) Toolkit, Lancaster University	Outlines the steps taken during the development of an evaluation plan. Offers ideas and materials to support evaluation planning and methods.	http://www.lancaster.ac.uk/fass/events/capacitybuilding/toolkit/
Web Center for Social Research Methods	Provides an overview of evaluation and how it differs from social research. Introduces several types of evaluation and the Planning-Evaluation Cycle.	www.socialresearchmethods.net/kb/evaluation.php

Annex 1: Guidance on outreach activity and evaluation type

There is no hard and fast rule as to what types of evaluation might be appropriate in different outreach contexts although the following generalisations may hold true in most cases.

Type of activity	Considerations	Suggestions	Best practice
Long term or multi- activity intervention Examples: Sustained progression with multiple interventions or compact type programme Intensive experiential activity Examples: Summer school or other residential programme	You can use quantitative or qualitative techniques or both. The approach should include repeated data collection with participants, Ideally, before or at the beginning, middle and endpoint as well as tracking into progression outcomes.	If the activity involves a large number of participants (more than 50) you will want to have some quantitative methods. Qualitative methods can complement the analysis. Using qualitative data creates a deeper understanding of how and why things work. If the programme includes different activities remember to record the take-up and frequency so that you can consider how differences in the interactions affect the outcomes.	Use both qualitative and quantitative methods to their strength. Use at least some of the same questions before, during and after the intervention to allow you to track participants' progress. Ensure that you can link individuals to their answers over time. Ideally, you will have a rationale for each of the questions you include and use validated questions wherever possible. For interventions such as a summer school or residential programme, you could try tracking those who have applied and were not selected for the programme as well as those who participated in the programme. This gives you a comparison group to gauge the effect of your activity (with the caveat that the groups are not completely comparable).
Intensive tailored support activities Examples: Mentoring, coaching, peer support from student ambassadors	You can use quantitative and qualitative techniques. Evaluation should include repeated data collection with participants. Ideally, before or at the beginning, middle and endpoint as well as tracking to progression outcomes.	You can also think about naturally occurring data, e.g. the content of mentoring or surveys as part of the delivery. Remember to consider differences by the frequency of interaction if this is recorded. Make sure you track the same individual throughout.	You could try to evaluate those who participated in the activities compared with those who did not participate (with their consent). This gives you a comparison group that gets you closer to singling out the effect of your intervention. If numbers are small, consider undertaking a qualitative evaluation.
Focused intervention with narrowly defined aims Examples: Academic tutoring, subject-specific enrichment activities	Depending on the objectives of the intervention and access to pupil data it may be possible to use objective measures (e.g. performance in exams), or you may need to develop your own tests in order to capture pre and post interventions measures of the outcomes (e.g. in terms of skills and knowledge developed by the participants).	As well as seeking to demonstrate a pre and post intervention change (e.g. in skills, understanding or achievement) you should also seek to demonstrate how the outcomes relate to HE progression outcomes. If you develop your own tests, remember to undertake cognitive testing, and ensure the results have internal validity.	A difference-in-difference design could be applied (i.e. looking at a pre and post intervention difference between an intervention and comparison group) however ideally you would need to make sure that the pre and post intervention measures were collected in the same way for the comparison as well as the outreach group).
One-off interventions	It may be hard to collect data more than once to balance the	A short survey that considers what participants have learnt	Include the academic staff who are giving their time for free in

Type of activity	Considerations	Suggestions	Best practice
Examples: Campus visits or open days, one-off subject taster sessions, one-off school visits	proportionality of time spent evaluating and doing an activity. For a Type 2 evaluation you may need to consider a comparison group (e.g. matched group of participants who did not take part).	on the day could be useful. You could also run a focus group inviting volunteers from several events to share their experiences – think about compensating them for their time. You may need to work closely with schools/colleges to make sure consents are in place.	the evaluation: you could ask them informally what they think the session achieved and how it could be improved. Equally, ask teachers in schools for feedback on how to make the activity even more valuable for the participants.
General information provision Examples: HE fair, careers fair	You want your evaluation to be proportionate to your activity. It is difficult to disentangle the effect of a one-off intervention on outcomes. Furthermore, the potential for following up participants tends to be limited because of a lack of access to individual participant information and consent.	It will be difficult to collect data and participants are busy taking in a lot of information. You want to keep distraction to a minimum. Feedback surveys are useful but keep questions very short, as respondents usually have little time on the day. Another approach is to use 'straw polls' for example using stickers to answer closed (yes/ no) questions about the outcomes.	You could collect contact details and have a survey after the event – perhaps with a prize draw. You could also have helpers undertaking very short semi-structured interviews with attendees, or undertaking follow-up (with permission).
Light-touch interventions Examples: HE-related careers information provision, master classes in schools	The potential for different types of evaluation will depend on the extent to which the provider has access to individual participants (for example for pre-intervention baselining) and to individual data (for example, consent to follow-up).	You will need to work closely with gatekeepers in schools and colleges to negotiate access to individuals and their data. Ideally you should seek to ensure that the delivery stakeholders are signed up to seeing the evaluation as an integral part of the intervention.	Depending on availability of participant details, it is possible to apply robust designs to light-touch interventions. For example, recent RCTs have been undertaken on a text-message based information campaign (see pages 33-34).

Glossary

Term (abbreviation)	Description	Link (if applicable)
Access Agreement	Until its closure at the end of March 2018, the Office for Fair Access (OFFA) required higher education providers to submit Access Agreements detailing their access and participation plans to be approved by the Director for Fair Access to Higher Education	
Access and Participation Plan (A&P Plan)	A regulatory requirement set by the Office for Students as the regulator for English higher education. Access and participation plans set out how higher education providers will improve equality of opportunity for underrepresented groups to access, succeed in and progress from higher education. They are approved by the Director for Fair Access and Participation	https://www.officeforstudents.org.uk/advice-and-guidance/promoting-equal-opportunities/access-and-participation-plans/
Administrative data	Collected as part of ongoing administration of education (e.g. school census, individual learner records, UCAS and HESA datasets, university databases).	
Aimhigher West Midlands	A partnership of universities, schools and colleges who work together to improve social mobility by delivering activities aimed at widening participation in all forms of higher education.	https://www.aimhigherwm.ac.uk
Appreciative inquiry (AI)	Model of planning and development that engages stakeholders in collective discussion about a policy, intervention or outcome.	
Before-after study	Research to assess impact involving obtaining data or measuring particular characteristics of a population before versus after an outreach intervention to measure the effect or correlation.	
Benchmarks	Performance data from previous or similar activity that helps you to set targets.	
Bivariate analysis	Type of statistical analysis that looks at two variables to determine the empirical relationship between them.	
Brightside	Third-sector organisation that helps young people make confident and informed decisions about their future through recruitment and training. Brightside mentoring is a social enterprise for online mentoring.	http://www.thebrightsidetrust.org
British Education Research Association (BERA)	Membership association and learned society committed to advancing research quality, building research capacity and fostering research engagement.	https://www.bera.ac.uk/
Case study	Aspects of a single case (such as a person, organisation or school/college) are investigated in depth within the case's own context. Case studies usually draw on a range of evidence in order to 'triangulate' the findings (i.e. drawing on different perspectives).	
Cohort study (sometimes known as a panel study)	Longitudinal study where people are followed over time to see whether differences occur among different groups.	
Collaborative activity	Work to improve access, success and progression that involves collaboration between multiple higher education providers and other organisations (e.g. schools, colleges, third-sector organisations)	
Comparison group	Generally comparable to your participants, used to demonstrate how results differ to those taking part in the outreach intervention. Usually comparison is between participants and non-participants	

Term (abbreviation)	Description	Link (if applicable)
	in outreach, although there could be more than two groups receiving different types of interventions.	
Contextual data	Data linked to higher education applicants denoting educational, social and economic background which is used to identify candidates whose success in existing qualifications may not reflect their potential to succeed in higher education	
Contextual information	Information contained within higher education application materials which is used as part of holistic assessment of candidates potential to succeed in higher education	
Comparison group	Group that does not take part in your activity against which the impact of participation can be measured. Comparison groups are selected to closely resemble the group that receives the intervention using population characteristics (e.g. if the intervention is held in a school, a class group that did not take part might be used as the comparison group if the young people are in the same age range and have similar characteristics and school experiences.	
Control group	Group that does not take part in your activity against which the impact of participation can be measured. Control groups are selected using randomisation or a scientific approach to ensure that the group of participants and non-participants are as similar as possible and distinguished as far as possible only by participation/non-participation in the intervention.	
Data controller	The person or persons who are responsible for determining how and for what purpose somebody's personal data can be used.	
Data processing	Any activity in relation to a dataset (including obtaining, recording or holding the data or carrying out analysis).	
Data Processing Agreement	A written contract which needs to be in place whenever a data processor is working with data provided by a data controller. The GDPR sets out what needs to be included in the contract. This agreement relates to a form of data sharing where a data controller shares data with another party that processes personal data on its behalf.	https://ico.org.uk/for-organisations/guide-to-data-protection/
Data processor	The person who processes personal data under agreement with the data controller.	
Data protection	Personal data must be held in strict confidence, held securely, and appropriate technical and organisational information security and processing procedures must be established and maintained to ensure that data is sufficiently protected against any unlawful or unauthorised use.	
Data sharing	Data sharing means the disclosure of data from one or more organisations to a third party organisation or organisations, or the sharing of data between different parts of an organisation.	
Data Sharing Agreement (sometimes known as Data Sharing Protocol)	A common set of rules to be adopted by the various organisations involved in a data sharing operation (which could form part of a contract between organisations). This is concerned with sharing personal data between data controllers – i.e. where both organisations determine the purposes for which and the manner in which the personal data is processed.	See ICO's code of practice on data sharing at: https://ico.org.uk/media/for-organisations/documents/1068/data_sharing_code_of_practice.pdf
Data subject	The individual whom the data is about.	

Term (abbreviation)	Description	Link (if applicable)
Department for Education (DfE)	Responsible for children's services and education, including higher and further education policy, in England.	https://www.gov.uk/government/organisations/department-for-education
Descriptive statistics	Using aggregates and percentages to describe the profile of a sample or population.	
Difference-in-difference (DiD)	Compares a before-and-after change in participants' prospects relative to that of non-participants (i.e. the change observed after versus before an intervention in two groups who are compared: one group that receives an intervention (experimental group) and a group that does not receive the intervention being tested (control group)).	
Disclosure and Barring Service (DBS)	The DBS administers and manages the disclosure service to provide a regulated 'one-stop' service for England and Wales offering access to records held by the police and in relevant cases information held by the DBS.	
Disclosure control	Reporting protocol to prevent individuals being identified in presentation of data tables. Standard disclosure control includes: Rounding all numbers to the nearest multiple of 5; Any number lower than 2.5 must be rounded to 0; Halves must be rounded upwards (e.g. 2.5 must be rounded to 5); Percentages based on fewer than 22.5 individuals must be suppressed; Averages based on 7 or fewer individuals must be suppressed.	
Disclosure Service	Part V of the Police Act 1997 makes provision for two different levels of criminal record checks. Once a check is complete, one of two Disclosures will be issued. The type of check carried out will depend on the nature of the position applied for. Both of these checks require a fee (but are free of charge for volunteers). **The Standard Disclosure** check contains: details of all convictions, cautions, reprimands and warnings held on the Police National Computer (PNC). A Standard check cannot reveal if a person is barred from working with children or vulnerable adults. **The Enhanced Disclosure** is the highest level of check available to anyone involved in regulated activity for a regulated activity provider working with children or vulnerable adults. Enhanced checks contain the same information as the Standard check but also include a check of the new barred lists and any relevant and proportionate information held by the local police forces.	
East Midlands Widening Participation Research and Evaluation Partnership (EMWPREP)	Formed in July 2011 as a partnership to utilise data to effectively target learners for widening participation interventions, capture and store learner data, and monitor and evaluate members' outreach work, with the aim of tracking learners through their educational lifecycle.	www.emwprep.ac.uk
Education Endowment Foundation (EEF)	The EEF and Sutton Trust are, together, the government-designated What Works Centre for Education, and makes available evidence-based resources to inform the practice of teachers and senior leaders.	https://educationendowmentfoundation.org.uk
Eligible expenditure	Financial costs that can be included within a higher education provider's agreement with the regulator on use of income to promote access to higher education.	

Term (abbreviation)	Description	Link (if applicable)
Evidence and Impact Exchange (EiX)	The EIX is designed to provide evidence on the impact of approaches to widening access and successful participation and progression for underrepresented groups of students, and will aim to ensure that the most effective approaches are recognised and shared (launch in spring 2019).	https://www.officeforstudents.org.uk/advice-and-guidance/promoting-equal-opportunities/using-evidence-and-evaluation-to-improve-outcomes/evidence-and-impact-exchange/
Evaluation	The process of making judgements on the success of what you are doing. There are various different types of evaluation: impact evaluation assesses the changes that can be attributed to an activity; process evaluation assesses the effectiveness of the processes involved in implementing the activity.	
Experimental design	Term used to describe research studies where one (or more) variable is being influenced whilst the researcher controls the other variable(s) to determine whether there is a causal relation between the variable in question and the outcome.	
External comparison group	People identified who have not been identified through your activities, e.g. other pupils who were non-participants.	
Fair access	Equality of opportunity for all those who have the potential to benefit from higher education, irrespective of their background, schooling or income, including to highly selective institutions.	
Financial support	Support to students given in the form of: bursaries and scholarships; fee waivers; 'in-kind' support, e.g. discounted accommodation or credit against spending on campus.	
Free School Meals (FSM)	Maintained schools and academies (including free schools) are required to provide free school meals to disadvantaged pupils who are aged between 5 and 16 years old. Eligibility criteria link to receipt of benefits. The use of FSM data is prevalent in educational research reports as a proxy for disadvantage, although the measure is likely to be a subset of those facing educational disadvantage.	https://www.gov.uk/government/publications/free-school-meals-guidance-for-schools-and-local-authorities
Fuzzy matching	The process by which data is linked at individual level from different sources when there is not a unique individual identifier. Most commonly data is linked using personal data fields such as name, postcode and date of birth.	
General Data Protection Regulation (GDPR)	European Union regulation to strengthen and unify data protection for individuals.	https://ico.org.uk/for-organisations/guide-to-data-protection/
HESA Student Record	The central source for the collection of statistics about students in publicly funded UK higher education. In general, the student record will be collected in respect of all students registered who follow courses that lead to the award of a qualification or provider credit. HESA is the data controller for the student record.	https://www.hesa.ac.uk
Higher Education Access Tracker (HEAT)	HEAT is a membership organisation formed in 2011 to help members to target, monitor and evaluate widening participation outreach programmes and to track students' progression from school into higher education and beyond.	www.heat.ac.uk
Higher Education courses	Programmes leading to qualifications (or credits towards) qualifications, which are above the standard of GCE A-levels or other Level 3 qualifications.	

Term (abbreviation)	Description	Link (if applicable)
Higher Education Funding Council for England (HEFCE)	Non-departmental public body responsible for the distribution of funding for higher education to universities and further education colleges in England since 1992. HEFCE closed at the end of March 2018	http://www.hefce.ac.uk
Higher Education Statistics Agency (HESA)	Responsible for collecting and publishing detailed information about the UK higher education sector.	https://www.hesa.ac.uk
IAG (Information, advice and guidance)	IAG refers to provision to make careers education more accessible for young people whatever their background, to support the right education and training choices.	https://www.gov.uk/government/publications/careers-strategy-making-the-most-of-everyones-skills-and-talents
IMD (The English Index of Multiple Deprivation)	Official measure of relative deprivation for small areas (or neighbourhoods) in England.	https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015
Income Deprivation Affecting Children Index (IDACI)	Part of the IMD, a measure of income deprivation among children.	https://www.gov.uk/government/statistics/english-indices-of-deprivation-2015
Individualised Learner Record (ILR)	A collection of statistical data returned at various points of the academic year by providers in the further education system.	https://www.gov.uk/government/collections/individualised-learner-record-ilr
Information Commissioner's Office (ICO)	The UK's independent authority set up to uphold information rights in the public interest, promoting openness by public bodies and data privacy for individuals.	https://ico.org.uk
Internal comparison group:	People you know who are similar to participants (e.g. if you have a competitive application process for your intervention because places are limited, you could use the group of rejected applicants as a comparison group (with their consent)).	
Intervention	What you do in order to bring about a change – i.e. the activities that you deliver to outreach participants.	
Linked data	Data linked to individuals from different sources, usually using a unique identifier or individual matching process. Most common sources of linked data are from the NPD and HESA student record.	
Longitudinal research	Research involving repeated measurements over long periods of time.	
Long-term outreach	Access work with learners that goes on over a sustained period of time and follows them through different stages/activities	
Low Participation Neighbourhood (LPN)	Quintile 1 and Quintile 2 areas in the POLAR classification of young HE participation.	https://www.officeforstudents.org.uk/data-and-analysis/polar-participation-of-local-areas/
Milestones	Short-term measures which track progress towards a long-term target.	
Mixed methods	Research which utilises both qualitative and quantitative research methods.	
Monitoring	The process of gathering and analysing information on your progress as you go along in order to see whether you are working as planned.	
Multiple Equality Measure (MEM)	Composite measure of disadvantage taking account of a range of socio-demographic factors being developed by UCAS.	https://www.ucas.com/corporate/data-and-analysis/ucas-undergraduate-releases/equality-and-entry-rates-data-explorer

Term (abbreviation)	Description	Link (if applicable)
Multivariate analysis	Type of statistical analysis that looks at more than one statistical outcome variable at a time (i.e. taking into account the effects of multiple variables).	
Narrative	A way of summarising your project in such a way that explains what you will do and why you will do it to bring about the change you want to see (i.e. telling the story about how long-term improvements will be achieved and why).	
National Collaborative Outreach Programme (NCOP)	Programme to support young people in some of the most disadvantaged areas in England. The programme targets young people in Years 9-13 who have the attainment level to progress into higher education. It runs from 2016-17 to 2019-20.	https://www.officeforstudents.org.uk/advice-and-guidance/promoting-equal-opportunities/national-collaborative-outreach-programme-ncop/
National Networks for Collaborative Outreach (NNCO)	Predecessor to NCOP, this programme brought together universities and further education colleges into local networks to provide coordinated access to schools and colleges.	
National Pupil Database (NPD)	Data collections from pupils in state-funded education in England. The Department for Education is the data controller for the NPD.	https://www.gov.uk/government/collections/national-pupil-database
NERUPI	Framework developed by the University of Bath which sets out defined aims and outcomes which are the key to effective evaluation.	http://nerupi.co.uk
Office for Fair Access (OFFA)	Non-departmental public body set up in 2004 to promote and safeguard fair access to higher education for underrepresented groups, in light of the introduction of variable tuition fees in 2006-07. OFFA closed at the end of March 2018.	www.offa.org.uk
Office for Students (OfS)	Regulator of English higher education providers with a remit for helping students to access, succeed in and progress from higher education; helping students stay informed; making sure that students get a high-quality education that prepares them for the future and protecting students' interests.	https://www.officeforstudents.org.uk/about/who-we-are/
Outcomes	The changes that you are aiming to bring about – such as how you will benefit your outreach participants including the learning/skills development and other benefits that happen as a result of taking part.	
Output	What you deliver in order to meet you outreach objectives – i.e. the number of activities, events, sessions, services, materials that you deliver.	
Outreach activity/Outreach work	Activity that involves raising aspirations and attainment among potential applicants from underrepresented groups and encouraging them to apply to and enter higher education. References to outreach work are designed to mean the outreach programmes, practices and innovations, policy developments, partnerships and collaborations and so on that take place in order to address underrepresentation in higher education.	
Pilot project	Often a small-scale trial of something that is designed to test whether a new idea or development works (e.g. running a new delivery mode with a small group of people first in order to find out if it is effective before rolling it out to a larger group).	

Term (abbreviation)	Description	Link (if applicable)
POLAR (Participation of Local Areas)	POLAR classifies local areas into five groups, based on the proportion of 18 year olds who enter higher education aged 18 or 19 years old. These groups range from Quintile 1 areas, with the lowest young participation (most disadvantaged), up to Quintile 5 areas with the highest rates (most advantaged). The latest version (POLAR4) is based on the combined participation rates of those who entered higher education between the academic years 2009-10 and 2013-14, if they entered aged 18, or between 2010-11 and 2014-15 if they entered aged 19.	https://www.officeforstudents.org.uk/data-and-analysis/polar-participation-of-local-areas/
Pre-condition	A factor which is needed for something to take place. For example, buy-in from teachers may be needed in order for participants to take part.	
Progression	Either a general term to denote movement through the student lifecycle or in the context of OfS guidance from graduation to work or postgraduate study.	
Propensity score matching (PSM)	Compares the outcomes of your participants with individuals who did not receive the intervention but who were as likely to have done so (i.e. aiming to correct for non-random selection).	
Qualitative research	Qualitative methodologies involve engaging with your participants through conversations (interviews, focus groups), observations or written form (e.g. open-ended questions on feedback questionnaires and surveys). Qualitative research usually aims to understand people's beliefs, experiences, attitudes, behaviour and interactions.	
Quantitative research	Quantitative evaluation methodologies include collecting new or using existing data in numerical counts from a representative sample or all your participants. This includes use of linked data to participants so you can draw new inferences on them, or collecting new data, for example, by doing a survey and analysing the results numerically.	
Randomisation	The process of selecting subjects entirely randomly, for example by drawing lots, which means that each participant (or other unit such as a school) has an equal chance of being in the intervention or control group in order that the influence of any distorting factors is equally spread and the groups are as comparable as possible (except for the outreach intervention).	
Randomised Control Trial (RCT)	An experiment to compare outcomes carried out with different groups where participants are randomly assigned to receive an intervention or not.	
Realistic evaluation	A model of theory-driven evaluation, which focused on finding what outcomes are produced from interventions, how they are produced and the conditions that make the intervention effective.	
Regression analysis	A type of inferential statistical modelling that estimates the relationship between an outcome variable and independent variable(s). Various regression models exist, and usually the aim is to predict the extent to which a variable changes the outcome variable when the other independent variables are constant.	
Regression Discontinuity Design (RDD)	Compares outcomes for individuals who are essentially equivalent but just fail to meet the criteria for participation on one key characteristic (e.g. people who have similar family income but may be above/below a threshold for participation).	
Reliability	Refers to consistency of results (i.e. the extent to which same tools would generate the same results each time used under same conditions).	

Term (abbreviation)	Description	Link (if applicable)
Response rate	Proportion taking part (e.g. in a survey), which is used to give an indication of how representative the results are. If you want to make generalisations, for example to say that results for a sample are the same as for a cohort in general, then you need to make sure the data is based on a relatively large group.	
Sensitive personal data	Data relating to a living individual that would allow the person to be individually identified (e.g. name, date of birth), plus other details that need to be kept private to prevent unwarranted disclosure (e.g. the person's race, ethnicity, politics, religion, trade union status, health, sexual orientation, criminal record, physical or mental health).	
Statistical significance	The probability that a finding is true and not down to chance. For a sample of 100 people a change of 12 per cent would be considered statistically significant (i.e. not down to chance).	
Stakeholders	The people or groups of people who are affected by your activities or who have an effect on it. These include the participants, schools and colleges and others such as parents/carers.	
STROBE	UCAS service that can track individuals into the UCAS applications system, and report anonymously on their outcomes or characteristics at aggregate levels. The cost of this service depends on the data requirements.	https://www.ucas.com/data-and-analysis/data-products-and-services/strobe
Target groups	Guidance on target groups varies from year to year. Target groups include those underrepresented in higher education such as: people from lower socio-economic groups or from neighbourhoods where higher education participation is low; people from low income backgrounds, some ethnic groups or sub-groups, including White males from economically disadvantaged backgrounds; disabled people; mature and part-time learners; care leavers; carers; people estranged from their families; people from Traveller communities; refugees; students with mental health problems, Specific Learning Difficulties and/or those on the autism spectrum; children from military families.	For details about a particular year, see the guidance for that year.
The Access Project (TAP)	A charity providing one-to-one tuition to students in disadvantaged areas.	http://www.theaccessproject.org.uk
The Sutton Trust	Foundation which improves social mobility in the UK through evidence-based programmes, research and policy advocacy.	https://www.suttontrust.com/
Theory of Change (ToC)	A coherent account of why your outreach intervention might have the effect you want and how your activities link to the desired results.	
Type 1 evaluation	A Type 1 evaluation provides a coherent account of why your outreach intervention might have the effect you want and how your activities link to the desired results. In order to meet the standard you will need to be able to refer to evidence of impact elsewhere and/or in the research literature on outreach effectiveness.	
Type 2 evaluation	A Type 2 evaluation provides data on impact and can report evidence that those receiving an intervention treatment have better outcomes, though this does not establish any direct causal effect. To meet this standard you will need quantitative and/or qualitative evidence of a pre/post treatment change or a treatment/non-treatment difference	

Term (abbreviation)	Description	Link (if applicable)
Type 3 evaluation	A Type 3 evaluation involves a methodology which provides evidence of a causal effect of an intervention. This type uses quantitative and/or qualitative evidence of a pre/post treatment change on a treated group relative to an appropriate control or comparison group.	
UK Performance Indicators in Higher Education (UKPIs)	The UK Performance Indicators provide comparative data on the performance of higher education providers across several areas including participation of widening participation students.	https://www.hesa.ac.uk/data-and-analysis/performance-indicators
Underrepresented groups	Groups that are currently underrepresented in higher education nationally.	
Univariate analysis	Statistical analysis where only one variable is being considered.	
Validity	Refers to suitability of the method used to answer a question (e.g. looking at exam results is helpful to measure education progress but might be less useful if you want to know about attitudes to higher education).	
Whole-institution approach	An approach to widening participation and fair access that is embedded at all levels of an institution, not limited to a particular unit or department, engaging across all areas of its work and inclusive of senior management.	
Widening participation	Removing the barriers to higher education, including financial barriers, which students from lower income and other underrepresented backgrounds face to progression to higher levels of education.	
Young student	Aged 20 and under on entry to higher education.	

---

[20] https://www.officeforstudents.org.uk/publications/understanding-the-evaluation-of-access-and-participation-outreach-interventions-for-under-16-year-olds/