Access and participation standards of evidence

This document lays out access and participation standards of evidence and discusses how higher education providers can strengthen their standards of evidence. It is for senior managers, decision-makers and practitioners with a remit for evaluation and reporting on access and participation activities. It is designed to provide guidance on what type of impact evaluation to aim for, ways to strengthen the evidence, and the claims that can be supported by different types of evidence. 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.

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

- Using standards of evidence to evaluate impact of outreach
- An evaluation self-assessment tool

1 Introduction

This guide lays out access and participation standards of evidence, which aim to support the work of the Office for Students (OfS) and higher education providers. It lays out how standards of evidence relate to types of impact evaluation and highlights what sort of claims are supported by different types of impact evaluation. Providers can use the standards to inform their impact evaluation strategies and access and participation activities and the OfS can use them to assess providers' claims of impact and in decisions about the evidence used to support policy and funding decisions.

What is evidence?

Evidence is the body of facts, results or data used to prove or support claims and assertions about impact and effectiveness. There are many different types of evidence, and the strength of evidence can vary. The OfS wants access and participation providers to use appropriate evidence for the types of interventions and where possible, aim for the strongest possible evidence.

Evidence is information to support assertions.

Different types of evidence are associated with different types of impact evaluations (and the aim of developing your evaluation practice is usually to ensure a higher standard of evidence and a more rigorous approach).

Different types of impact evaluation generate different evidence.

What is evidence for?

Evidence supports the development of conclusions and recommendations about activities, and learning about how to improve practices, to better support under-represented groups to access and succeed in higher education. Evidence is used to guide investments and inform the design and delivery of access and participation activities.

Evidence supports claims and underpins design and delivery of programmes.

Why are standards of evidence important?

Evaluation of the impact of investment in access and participation involves making judgements about the results of the activities, based on an assessment of the available evidence. The two main aims are accountability for the results achieved by access and participation interventions and learning whether the activities work or not. These standards provide a framework for looking at evidence of impact (rather than for other types of evaluation or research). Using a shared reference framework of standards for evidence of impact facilitates knowledge sharing and making judgements and can thus increase the effective use of evidence by practitioners. As providers develop stronger and more robust impact evaluation plans, higher quality evidence will be generated. Ultimately, by sharing findings about what is expected to work in what circumstances with whom, what is proven to work and what does not work, the use of standards will help to ensure that access and participation activities and funds are directed to the most effective activities.

Evidence is important for accountability and learning.

The ultimate aim is to strengthen evaluation of interventions in order to ensure resources are deployed most effectively.

How should providers use standards?

Higher education providers should use standards of evidence:

- in deciding what evidence to use to guide their decision about whether to invest in different types of interventions and practices, and to help them to improve their access and participation delivery and performance;
- to guide what kind of evidence is generated by their own impact evaluation and analysis and to clarify the claims that can be made when reporting the results of evaluations externally.

Use standards to decide which evidence informs delivery and practices, to decide how to evaluate and the claims that can be made from impact evaluations.

The standards framework

The standards aim to facilitate robust and rigorous impact evaluation by providing a shared reference framework across higher education providers. In line with evaluation tools and standards developed elsewhere, the model is based on three types of evaluation which generate different types of evidence: the **narrative** of evaluation

Three types of impact evaluation: narrative;

(knowing what you are doing and why); **empirical enquiry** (evaluation to measure changes generated by different activities and practices); and **consideration of causal claims** (what impact can you identify as a direct result of your activities?).

empirical; and causal.

The three types of impact evaluation are distinct but not hierarchical. The standards are concerned with the robustness and certainty of the evaluation strategy and appropriateness to the activities you are undertaking. It is possible to make different judgements about the impact of access and participation activities from the different types, so long as the evidence from the evaluation is robust, high-quality evidence.

No hierarchy is implied in the different types. What matters is the quality of the evidence.

	Description	Evidence	Claims you can make
Type 1: Narrative	The impact evaluation provides a narrative or a coherent theory of change to motivate its selection of activities in the context of a coherent strategy	Evidence of impact elsewhere and/or in the research literature on access and participation activity 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 Enquiry	The impact evaluation collects data on impact and reports evidence that those receiving an intervention have better outcomes, though does not establish any direct causal effect	Quantitative and/or qualitative evidence of a pre/post intervention change or a difference compared to what might otherwise have happened	We can demonstrate that our interventions are associated with beneficial results.
Type 3: Causality	The impact evaluation methodology provides evidence of a causal effect of an intervention	Quantitative and/or qualitative evidence of a pre/post treatment change on participants relative to an appropriate control or comparison group who did not take part in the intervention	We believe our intervention causes improvement and can demonstrate the difference using a control or comparison group

The standards of evidence build on the work of Crawford et al. in 2017 that established a framework for types of evaluation of the impact of outreach (commissioned by the Office for Fair Access (OFFA) and the Sutton Trust).

It is expected that all types of impact evaluation findings will feed into the access and participation decision-making processes. Impact evaluation can inform decisions on the design or implementation of an intervention, decisions on whether to continue or stop an activity, and strategic decisions on a certain type of practice or delivery aspect.

Evaluation informs decision making.

There is an element of **proportionality**; the impact evaluation methodology used should be proportionate and appropriate to the type of activity and its stage of development. What is required might depend on the existing evidence base because if the activity is a new innovation or as yet unproven then your evaluation strategy should collect empirical or causal evidence of its results. If previous reliable evaluations have shown that a particular activity gets good results in context then there is less need to implement impact evaluation to test effectiveness.

Choice of type of impact evaluation depends on the activity and type of evidence needed to demonstrate whether the activity is effective.

Good evidence is not confined to particular data or methods. What is important is **high-quality, robust and appropriate evidence.** Such evidence identifies the most promising new approaches and in reviewing existing interventions to ensure they continue to work.

Evidence needs to be high quality in order to allow a judgement.

Evidence from impact evaluation could be based on more or less robust and developed practice. **Depending on the strengths and limitations of the approach we can be more or less confident in the accuracy of the results in providing a true reflection of an activity's impact**. Usually the level of rigour at the planning stage, the extent to which the evaluation is systematically implemented, and the reliability of the approaches to data and analysis will affect the robustness of the evidence that is presented by the evaluation.

Robust evidence requires rigour in the planning stage, systematically implemented evaluation and reliable data and analysis.

There is a **link between evidence and practice.** It is generally not possible to have strong evidence where actual practice is weak. Stronger practice draws on and creates strong evidence. Therefore, to an extent, indicators of evidence are also indicators of practice.

Overview of Types of Impact Evaluation:

Type 1: Narrative		**Type 2: Empirical Enquiry** *(encompasses Type 1 and the following)*		**Type 3: Causal claims** *(encompasses Type 2 and the following)*	
✓ **Yes Please**	✗ **No thanks**	✓ **Yes Please**	✗ **No thanks**	✓ **Yes Please**	✗ **No thanks**
Coherent strategy	Disjointed activities	Clear aim of what activities seek to achieve	Aims developed after activity	Have a target as well as a control or comparison group	
Approach and activities underpinned by evidence from literature or other evaluations	No rationale for developing approach and activities	Select indicators of your impact	No concept of measuring success	Could use an experimental or quasi-experimental design	Using groups that are not comparable
Shared understanding of processes involved	The model of change is not shared	Quantitative or qualitative data – or both, 'triangulation' is good!	Information not systematically collected	Think about selection bias and try to avoid it	Selection bias in comparator groups
Reason for activity	Ad hoc activities	Pre/post data (minimum two points in time)	Only collect information once		
Clear conception of why the changes you seek to make are important	No understanding of needs of target groups	Analysis competently undertaken	Data not related to the intervention		
Programme reviews	No review or evaluation	Sharing of results and review of activity	Results not used to inform decisions		

2 Guidance on achieving a good standard of evidence

2.1 Getting ready for a Type 1 evaluation

A Type 1 evaluation provides a coherent account of why your access and participation intervention might be effective and how your activities link to the desired results. In order to meet the standard you'll need to be able to refer to evidence of impact elsewhere and/or in the research literature on effectiveness.

For Type 1 you'll refer to existing evidence of the impact of the activities.

General impact evaluation questions that are not overly specific to the intervention in question might be answerable via a qualitative review (or a more formal systematic analysis) of the existing literature. Review-based methodologies will be especially useful where there is already convincing evidence pointing to results that are transferable to different contexts, enabling your impact evaluations to focus more closely on the specific questions which the current evidence base leaves unanswered.

A review of the literature can be useful where there is a body of existing evidence to draw on.

2.1.1 Evidencing a Type 1 evaluation

There is a range of evidence that can support the requirements of a Type 1 evaluation. This list is indicative rather than exhaustive.

Dimension 1 of Type 1 evaluation. An evidence base for the activity or activities being undertaken, either referring to existing evidence of impact which you or others have collected and/or in the research literature on effectiveness. This should contain some or all of the following:

- Citations and references to relevant theoretical or practitioner literature, including scholarly literature as well as government and other reports, and explaining how these feed into your own practices.
- Participation in conferences or other types of engagement and knowledge sharing with other practitioners at regional, national or international level, with evidence of how this knowledge exchange feeds into your own practices.
- Evidence of keeping continuously up-to-date, including review cycles for renewing literature reviews, with evidence of how the knowledge is used in improving practice.

You can point to relevant literature/reports, knowledge exchange activities, or reviews and show how these have informed your activities.

Dimension 2 of Type 1 evaluation. An underpinning intervention logic (sometimes called theory of change or logical framework model) which comprehensively describes how and why a desired change is expected to happen in your particular context.

- You could have one overarching model for your access and participation programme, so long as there is clear linkage showing how activities are underpinned by clarity on the process by which interventions are expected to generate positive improvements.
- You could have individual models for different activities (linked to an overall coherent strategy for access and participation).
- The model(s) should be based on defined outcomes of each activity and the overall widening participation strategy.
- You should be able to demonstrate how the intervention logic was developed in conjunction with others beyond the widening participation team e.g. through informal practices, meetings, seminars and/or committees in order to draw on the knowledge and opportunities for comment from partners and stakeholders.
- As for possible gaps in knowledge, the complexity of concepts such as 'aspiration', and the complexity of young people's decision making around higher education should be acknowledged in the intervention logic.

You have a well-developed model comprehensively describing how and why a desired change is expected to happen in your particular context, informed by the literature and the knowledge and experience of your internal and external stakeholders.

Evidence supporting a Type 1 evaluation is thus:

- ***An evidence base for what you are doing, either referring to existing evidence of impact which you or others have collected and/or in the research literature on effectiveness.***
- ***A well-articulated intervention logic for your programme theory (such as a theory of change or logical framework) which describes how your activities will lead to the outcomes you hope to achieve, and the processes involved in bringing about the positive improvements.***

Type 1 evidence is strengthened by having a grounding in the existing evidence base on the impact of access and participation activities, an underpinning rationale for what you want to achieve and why, and coherent activities competently delivered.

2.2 Getting ready for a Type 2 evaluation

A Type 2 evaluation collects data on impact and can report evidence that those receiving an intervention have better outcomes. To meet this standard you will need quantitative and/or qualitative evidence of a pre/post change (i.e. an improvement after taking part compared to before the activity) or a treatment/non-treatment difference (i.e. an improvement over what would have happened without the activity or above what others achieve).

For Type 2 you'll need to demonstrate a difference in outcomes compared to what otherwise might have happened.

It is possible to choose from a range of methodologies, and some impact 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 the claims you want to make, then select a design that will enable you to achieve your aims for the evaluation.

Different approaches to the collection of data are possible.

Quantitative evaluation methodologies include collecting new or using existing data in numerical counts from a representative sample or from all participants. This could include use of secondary data (e.g. data on exam results or higher education (HE) applications) to draw inferences, collecting new data, for example, by doing a survey and analysing the results numerically. If you are working with many students, you might want to undertake quantitative methods using statistical analysis tools to compare between groups. This can be useful for not only gauging effectiveness in improving outcomes such as attainment and HE participation etc. for your participants but also in making inferences about how the outcomes of the activities could be generalised to wider cohorts.

Quantitative evaluation methodologies which measure the outcomes against a counterfactual can be useful for making inferences about the benefits for the target groups.

Qualitative evaluation methodologies are useful for gathering perspectives of the outcomes and the processes involved in achieving the desired result. Qualitative data can be collected using a variety of tools such as interviews, focus groups, case studies, artefacts, capturing personal experiences through visual texts, or direct observations. If robust data analysis is in place, qualitative data that is collected systematically can often be used to generate quantitative data. If your intervention is in the pilot stage you might want to use qualitative data such as interviews or focus groups to understand the processes involved which achieve the best outcomes and impact.

Qualitative evaluation methodologies can be useful in understanding the processes involved in generating outcomes.

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

2.2.1 Evidencing a Type 2 evaluation

In designing a Type 2 evaluation it is useful to think about what different types of evidence methodologies might be most useful in i) establishing the relationship between your activity, the objectives you are aiming for and the observed impact, ii)

Chose a method that establishes the outcomes and enhances

enhancing your understanding of the strength of any effect and the plausibility of your programme theory, and iii) coherence with other existing evidence about the activity.

understanding of the effects of the intervention.

Especially for strong Type 2 and for all Type 3 evaluations, you want to have a counterfactual or comparator to establish the impact of your intervention or activity above what might otherwise have occurred. For example, by measuring changes for a comparator group that did not take part in the activity but is as similar as possible to the intervention group, you build the case that your activity is what makes a difference to outcomes. Setting up a suitable comparison group is crucial where the impact evaluation is seeking to explain outcomes for some young people compared to a cohort who did not receive the intervention (e.g. those in a school cohort), or compared to the wider population (e.g. similar young people nationally). Having a comparator group is not always easy or straightforward to accomplish, but it is also not impossible. Further guidance is given in the 'Using standards of evidence to evaluate impact of outreach'.

Having a comparator group is one way of showing the difference for your participants.

If you are collecting data on your participants, ideally you want to aim for individual-pupil-level data rather than aggregated measures, such as whole cohorts. Individual data allows you to track how one particular 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. Having individual-level data is particularly important when tracking individuals across the student lifecycle.

Using individual-level data is better than aggregate (group/cohort) level data.

Unless you have a true control group – for example if you are doing an experiment such as a randomised controlled trial (RCT) – establishing causality between your access and participation activities and different outcomes may involve complex analysis. It may help to *triangulate* your findings i.e. to implement evaluation and data collection from different sources to enable the analysis to draw on a range of perspectives. You can also use findings from the statistical analysis alongside other qualitative or quantitative evaluation methods. An approach with draws on different methods involving both quantitative and qualitative analysis (sometimes referred to as a 'mixed method' approach) can also help you to identify the consistency of the evidence with the intervention logic (i.e. what you were expecting to happen) and where necessary to identify and explain the conditions under which the intervention is seen to operate. Mixed-methods evaluations draw in evidence from different sources and you will need to make sure the evidence is analysed systematically in order to present reliable evaluation results. 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.

It is often useful to employ mixed-methods research (i.e. using different techniques) in order to take account of different perspectives on the outcomes.

The results from interventions can change over time depending on the context in which the work is taking place. For example you could observe a trend towards higher rates of HE entry for participants above that of their peers but this improvement could be due to a range of factors. Therefore you usually need other research with learners and teachers for example to assess the relative strength of the intervention effect particularly if the sample sizes are small. Drawing in perspectives from teachers and other stakeholders such as parents and carers can be especially useful when evaluating interventions for younger age groups since the measures of success and the relationship with HE progression are less well understood and harder to capture than for older age groups who are further along their journey towards HE.

Stakeholders such as teachers and parents can provide useful insights, as well as the participants themselves.

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 a question at different levels and in more depth. A staged approach can be taken to interpretation – 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 quantitative data).

Quantitative and qualitative methods are complementary.

Evidence supporting a Type 2 evaluation is thus:

- ***Able to demonstrate a change above and beyond what might otherwise have been reasonably expected to have occurred.***
- ***Drawn from different research traditions and evaluation approaches (encompassing quantitative and/or qualitative methods).***

Different practices are associated with weaker and stronger evaluation evidence. Type 2 evidence can be strengthened by using appropriate indicators of impact, capturing changes over time, using valid data collection tools, robust sampling, an appropriate analytical strategy, and recognition of any limitations of the approach.

2.3 Getting ready for a Type 3 evaluation

The difference between a Type 2 and Type 3 evaluation is the level of confidence with which the impact observed can be attributed to 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 including experimental or quasi-experimental design.

Type 3 establishes that the intervention caused the outcome.

2.3.1 Evidencing a Type 3 evaluation

This type uses quantitative and/or qualitative evidence of a pre/post treatment change on the group taking part in the activity relative to an appropriate control or comparison group. Establishing the effectiveness of an intervention requires use of specific research designs, which are generally accepted as offering robust experimental and quasi-experimental designs capable of indicating causal effects. Evaluations utilising longitudinal tracking, regression discontinuity design (RDD), and RCTs are some of the stronger designs. RCTs have the highest evidential results, although other designs also generate strong evidence – for example using matched groups based on relevant characteristics (such as socio-demographic and educational variables).

Specific research designs need to be used which are capable of establishing causality.

Of course, as well as having an appropriate design, the implementation also needs to be robust and rigorous not least to minimise any spurious results. Type 3 evaluations are notoriously challenging because of the requirements for appropriate samples, for ensuring that appropriate data can be obtained and the appropriate consents to use the data put in place, and for complicated statistical analysis. Type 3 evaluations also tend to be costly. This type will not be proportionate or feasible for many activities.

Type 3 evaluations depend on data availability and can be resource intensive. This type will not be proportionate or feasible for many activities

Evidence supporting a Type 3 evaluation is thus:

- ***A research design methodology that establishes the extent to which observed results are caused by an intervention.***

Type 3 evidence is strengthened by using appropriate outcome measures, a robust research design, and an appropriate analytical strategy recognising the significance and strength of any effects.

Key Terms

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 a statistical technique or a ‘natural’ experiment is used to build a comparison group as similar as possible to the intervention group in terms of pre-intervention characteristics and conditions.

Review and Reflect: Two worked examples of different types of evaluations

EXAMPLE 1: Information through a football club	EXAMPLE 2: Individualised tuition for disadvantaged students
Overview: The approach brings current students from a partnership of universities into a local football club to help with the coaching of the players and, at the same time, to deliver information, advice and guidance (IAG) about HE. These students are either studying sports science courses or are members of university football teams. Other activities such as university visits have been organised to reinforce the student coaches' work. **About the evaluation:** The research focused on the benefit of participation to their school performance and to their HE aspirations and awareness, their learning from the project and intention to consider HE study. **Methodology:** The opinions of participants about the project were surveyed by questionnaires and focus groups at the end of the project. The evaluation used a questionnaire that had been tested in the previous year to provide comparative date over time. Research with the student coaches and parents/carers complemented the project. Over to you: What type of impact evaluation do you think this intervention is? Why? What is good about this evaluation? How could the approach be improved? **Possible Answers:** What type of impact evaluation is this? Type 1 Why? The partnership gathered data that shows some change amongst those receiving the intervention. It is developing a narrative account to motivate its selection of outreach activities in the context of a coherent outreach strategy. **What is good about the evaluation?** There is a holistic approach to understanding change for the participants in their context of sport and home. **How could the approach be improved?** The evaluation is based on research carried out shortly after the completion of the project, so only provides an assessment of views of the short-term impact. The introduction of longitudinal tracking is desirable to show the medium- and long-term impacts such as variations in attainment and HE progression rates. Comparative research, pre/post intervention, or gathering data on outcomes from a matched sample of participants and non-participants would help to confirm that the work is making a difference.	**Overview:** The project supports young people from disadvantaged backgrounds to progress to selective universities through offering one-to-one academic tuition in schools with volunteer tutors and personalised university support and guidance. **About the evaluation:** The evaluation aimed to assess the effectiveness of the tutoring programme at raising pupils' GCSE grades. **Methodology:** For GCSE grades – a matched comparison group design using propensity score matching and pupil data. For university places – pre/post data showing the change in number of pupils attending top universities from each school, from before the project started working with them to after. Over to you: What type of impact evaluation do you think this intervention is? Why? What is good about this evaluation? How could the approach be improved? **Possible Answers:** What type of impact evaluation is this? GCSE outcomes are Type 3 and university places are Type 2. Why? GCSE outcomes: They can demonstrate causality using a control or comparison group. Propensity-score-matching is widely considered a robust approach to creating a comparison group, provided that the factors on which participants are matched are sufficiently comprehensive and meaningful. (N.B. The project was unable to include 'level of motivation' as a matching factor, but was able to provide evidence to successfully make the case that this does not significantly weaken the findings). University places: The evaluation design compares the outcomes of pupils in the project with pupils from the same school who did not take part in the previous period – this gives an interesting benchmark, but there is likely to be some systematic difference between those pupils who did and did not take part in different years. **What is good about the evaluation?** There are robust 'before' and 'after' measures for the intervention. GCSE grades are externally verified measures. **How could the approach be improved?** By introducing a matching technique that mitigates bias, they can confirm with more certainty and accuracy that the programme is having a positive impact. For future impact evaluations this was tackled by using UCAS Strobe data to compare the participant outcomes against a matched 'control' group. Using focus groups or interviews could highlight which aspect of the tutoring is particularly helpful.

3 What can you say from your impact evaluation?

3.1 Claims you can make

The different types of impact evaluation and evidence provide increasing levels of surety about what makes a difference to learners and students in accessing and participating in HE. So before you decide on what type of evaluation you are going to undertake it's a good idea to think about what you want to be able to say from your evaluation findings.

The different types of impact evaluation affect the claims you can make about your impact.

As a general rule:

- Type 1: Essential to presenting a plausible rationale for why you are doing what you do.
- Type 2: 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 continuing your activity is worthwhile (without establishing definitive direct causal effects).
- Type 3: Important to use if you think an intervention is going to be effective but you need to have a high level of assurance that it works and need to be confident in the evaluation 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.

The types are all important for learning about what works and to improve effectiveness.

Usually, unless you have a very rigorous Type 3 approach, the relationship between the outcomes observed and an intervention can only be inferred. Nevertheless your results will still be important to contribute to learning about the effectiveness of access and participation interventions and highlighting the areas needed for improvement or reconsideration of an intervention. The improvements can then be made in the next round of delivery.

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

The conclusions you make must be informed by the results.

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 find that recipients have significantly better outcomes than the comparison group (suggesting either that your activity is effective at improving outcomes or you are seeing the effect of an unobserved bias in your intervention sample). If recipients have the same outcomes as the comparison group, this suggests 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. If recipients have significantly worse outcomes than the comparison group your activity must be considered detrimental for the target group objectives.

3.2 What do I do if my evaluation shows that my intervention does not have any impact?

Evaluation can sometimes be seen as an important source of validation for the activities being delivered and this poses a dilemma for practitioners and evaluators who may feel

that the public relations motive for an evaluation compromises the transparency of the evaluation. For example an evaluation might show that there is no evidence of participants having an increased chance of entering HE even though they took part in an activity that was well received at the time. However it is crucially important that evaluation is not subordinated to political or public relations purposes. Evaluation should be driven by practitioners' sense of responsibility to contribute to what is known about the outcomes of activities whether this is supportive of an intervention continuing or otherwise. Doing evaluation well matters because it helps to identify where interventions are not having the desired effect and therefore can prevent widening participation monies being wasted on ineffective interventions. Evaluation findings can show the relative utility of different interventions and enable funding to be prioritised accordingly. In this respect evaluation promotes accountability for spending.

Learning about what is not effective is just as important as finding out what works because both contribute to making sure resource is directed appropriately.

Evaluation activity is about both proving and improving, which means highlighting where impact was not achieved yet or the ambitions were unrealistic. Widening participation practitioners are working toward improving the effectiveness of access and participation programmes and evaluation is an important means to achieving this end. Impact evaluation helps providers to do their work better (helping not only to improve the effectiveness of the evaluated intervention but also helping to improve practice and analysis across the sector). 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.

Evaluation can support continual improvement in delivery.

Impact evaluation should account for both the positive and intended impacts and also any unintended or negative impacts of an intervention. Once an evaluation has attributed impact or lack of impact to a specific activity it often concludes with a judgement about the intervention's overall success or failure. The OfS recognises that practitioners are often reticent to admit 'failure' for fear of consequences. However it is important that stakeholders have the freedom to acknowledge any failure. The widening participation community is still learning about what works best in terms of impact and is moving away from absolute concepts of success and failure and instead recognising degrees of success and failure, which could depend on the target groups and context. Understanding the relative merits of different approaches in context is important to maximise the opportunities for learning.

The OfS encourages sharing of understanding of what has not worked so well and recognises that we need to understand more about the relative merits of different types of activities for different groups.

Your impact evaluation might show that an intervention works for some groups rather than others or only under certain conditions (and other evaluation may suggest explanations for this). These conclusions might not be robust findings in their own right but lead to new hypotheses which will need further testing to verify them. It is useful to capture and document the emerging hypotheses (for example as changes to the original intervention logic model) distinguishing between conclusions that are supported by evidence and new hypotheses for further testing.

Evaluation can identify further questions that need to be tested about what works for whom in what circumstances.

3.3 Sharing results of evaluations

Sometimes evaluations are designed and conducted without an explicit understanding of who is going to see or use the evaluation results. However given the focus on transparency and accountability for funding it is ethically difficult to justify evaluation if it will not be shared with others. It is important that widening participation practitioners agree to increase sharing of evaluation results (even where competition within the field militates against this).

Your results should be shared with the wider widening participation community.

4 When to use different types of evaluation

4.1 What questions are you seeking to answer?

When undertaking an impact evaluation access and participation practitioners are usually looking for the answer to this question: Does the activity make a difference to the HE access and progression outcomes that would otherwise have happened? (i.e. Was there any improvement in people's HE prospects?). There are many equally important questions (such as: Is this the right intervention for these people? Which intervention gets the best outcomes in which context? What makes the intervention successful?). Consideration of all of these questions may be beneficial at different times and will influence your choice of evaluation approach.

Your approach will reflect the specific questions the evaluation seeks to address.

Many access and participation programmes are complex and multifaceted. There may therefore be a need to unpack the relevant components of complex interventions in order to model multiple causal linkages and influences and thus gain a better understanding of how a programme works. Using a mix of research methods within your evaluation framework can be very useful in answering a range of research questions and providing a more rounded picture of what is going on.

Mixed-methods evaluations are useful for understanding processes as well as capturing outcomes.

4.2 What standard of evidence should I aim for?

The selection of evaluation approach is important since higher quality research designs can help to meet the challenge of attributing outcomes to the activity in question (as opposed to other influences) whereas lower quality designs reduce confidence in whether it was the activity that generated the outcomes. However there is no simple answer to the question of what will provide the best evidence for any particular type of access and participation activity. It depends on what is being measured and in what context.

Higher quality evidence is needed for more resource-intensive activities in order to test whether it is worth spending the level of resource.

As a rule of thumb the more resource-intensive an activity the higher one would wish the standards of evidence to be to show impact because it would be risky to continue to devote the level of resource unless the activity can be shown to have the beneficial impact it is aiming for.

What type of impact evaluation fits well with different categories of activities?

- Type 1 (Narrative) is expected as a minimum for all types of access and participation activities;
- Type 2 (Empirical) is expected for long-term or multi-activity interventions including mentoring schemes, and resource intensive activities, such as summer schools or other HE-residential programmes. This type of evaluation which is based on capturing indicators of impact over time is not expected in the case of very 'light-touch' activities, for example, one-off information provision or an ad hoc master class, campus visit or open day, or HE fair;
- Type 3 (Causal) is not an expectation, however it is recommended for certain types of costly interventions, and innovative and pilot interventions, so long as the expertise and resources are in place to develop and manage an experimental or quasi-experimental design.

The standard of evaluation expected depends on the type and nature of the activity.

The type of evaluation should be proportionate to the type of activity and scale of delivery.

These suggestions are intended to be illustrative only as there will be huge variation in practice in the nature of each of the activities being delivered. For example if relatively few students are involved in a mentoring activity and/or if it is very light-touch mentoring then it is unlikely that a Type 3 standard of evidence is appropriate. As the

‘owners’ of the evaluation it is for the individual higher education provider to determine the evaluation effort across different access and participation activities.

The matrix below sets out an indication as to which types of evaluation might be appropriate for different types of activity, though this should not be regarded as restrictive and, in particular, will vary according to the nature of the project and objectives.

	Narrative Evaluation *A coherent explanation of why the activity is expected to work and why*	**Empirical Enquiry** *Collects data to show those receiving an intervention have better outcomes*	**Establishing Causality** *Provides evidence of a causal effect*
Multi-activity intervention programmes (e.g. transition support programmes)	◆ Important for all activities to inform programme choice and delivery	◆ Important for all activities to justify use of resources	✪ Commended for resource intensive programmes for which an evidence base needs to be established and where there is access to reliable outcomes data and appropriate comparison groups
Intensive interventions (e.g. residential programmes)	◆ Important for all activities to inform programme choice and delivery	◆ Important for all activities to justify use of resources	✪ Commended for resource intensive programmes for which an evidence base needs to be established and where there is access to reliable outcomes data and appropriate comparison groups
Long term interventions (e.g. mentoring programmes)	◆ Important for all activities to inform programme choice and delivery	◆ Important for all activities to justify use of resources	✪ Commended for resource intensive programmes for which an evidence base needs to be established and where there is access to reliable outcomes data and appropriate comparison groups
One-off interventions (e.g. campus visits, subject taster sessions)	◆ Important for all activities to inform programme choice and delivery	★ Important for innovative projects for which an evidence base needs to be established	⃠ Not usually feasible unless part of a multi-intervention package
'Light-touch' intervention (e.g. information dissemination projects)	◆ Important for all activities to inform programme choice and delivery	★ Important for innovative projects for which an evidence base needs to be established	✪ Commended in situations where it is possible to capture appropriate outcomes data and the effect of the intervention can be isolated

◆ Expected for all types of activities; ★Commended for resource intensive and pilot interventions; ✪Highly commended if conditions allow and conducted appropriately; ⃠ May not be feasible unless special conditions apply.

The types of evaluation are not hierarchical – i.e. it is not a matter of trying to aim for a 'higher' type. Indeed it is better to aim for a strong Type 2 evaluation as opposed to an unrealistic or badly executed Type 3 evaluation. A well-formulated Type 2 evaluation is often more realistic, especially in the context of outreach where there are good working relationships in place with stakeholders who can provide access to good quality data and insights. Indeed establishing causality is likely to be complex and many RCTs have shown no effect, especially for interventions which target large groups of participants, in the case of particularly complex projects and programmes, or if there are significant confounding factors (which could include a wide range of influences on students).

Type 3 evaluations are not appropriate for all activities. A well-formed Type 2 evaluation is often more realistic.

4.3 What does the existing evidence show?

To an extent your choice of evaluation may be affected by what conclusions can be drawn from the existing evaluation results. If there is already convincing evidence that the activity is effective in causing the results desired then you may not need an experimental design bearing in mind that there is currently a dearth of rigorous, high-quality evidence on the impact of access and participation activities. However factors such as the characteristics of your target groups, phase of education and the environment/context in which you are working are going to affect whether the conclusions of previous studies are transferable to your particular situation.

Your approach should take account of what is already known about the effectiveness of the activity in question, and aim to fill gaps in knowledge.

5 Closing remarks

As the emphasis on managing widening participation resources for results increases, the importance of rigorous and evidence-based evaluations is mounting. It is in all providers' interests to undertake as high quality evaluation as possible as this will enable you to demonstrate your results and contribute to the understanding of what works in widening access and participation in HE. The standards of evaluation and standards of evidence are designed to help you to improve the quality and consistency of evaluations of the impact of access and participation interventions, whilst making sure that evaluation is appropriate and proportionate to the investment and activities.

We hope that this guidance will help you to develop increasingly rigorous and systematic approaches to evaluting the impact of access and participation activities, to act on the learning to achieve the best results for widening participation learners and students, and to contribute to the knowledge and understanding across the sector. Whatever your starting point, actions are possible to ensure increasingly robust evaluation and evidence which will contribute to the reduction of disparities in HE access and the strengthening of effectiveness of access and widening participation activities.

The OfS is seeking to develop and strengthen support on evaluation to providers over time and welcomes your comments and feedback on these materials.

5.1 Where can I find out more about evaluation of impact?

Using standards of evidence to evaluate impact of outreach
The guidance is for people who already have some experience with evaluation techniques and are looking to make impact evaluations more robust and embedded. The document highlights practices that can strengthen the impact evaluation of outreach, and offers case studies and signposting to further sources.
See: https://www.officeforstudents.org.uk/publications/standards-of-evidence-and-evaluating-impact-of-outreach/

An evaluation self-assessment tool
Self-assessment involves reflecting on your approach to impact evaluation against a series of questions. This tool has been developed to assist providers to review whether their impact evaluation plans and methodologies go far enough to generate high quality evidence about the impact of their activities in the Access and

Participation 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, and sets out the Standards of Evaluation of the impact of outreach, guidance on the standards and worked examples.

Acknowledgements:

This guidance was produced as part of a collaborative project led by the Centre for Social Mobility at the University of Exeter. We would like to thank the institutional and third-sector partners who kindly supported the work: Loughborough University, University of Liverpool, University of Plymouth, Royal Northern College of Music, The Access Project, Brightside, Coachbright, The Sutton Trust; and to the Office for Fair Access (OFFA) and The Sutton Trust who commissioned the project. The team are grateful for the support of the OfS who commissioned the project, and steering group members who helped to shape the outputs. We would like to thank Neil Harrison, Katy Vigurs, Julian Crockford, Colin McCaig, Ruth Squire and Lewis Clark who undertook research to understand the evaluation of access and participation outreach interventions for under 16 year olds,[1] whose work fed into this guidance.

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