what works centre for
local economic growth

Better impact evaluation in 9-steps

This series presents nine steps to make impact evaluation better, simpler, and less expensive. It provides key information for policymakers to refer to when thinking about impact evaluation of their local programmes.

This document brings together the blog series on how impact evaluation should be designed, originally published in spring 2015. We intend to demystify the process of undertaking robust and useful evaluation of local growth programmes and encourage more policymakers to undertake such evaluations. We have added a ninth step – Get help – which was not included in the original blog series.

1. Start early

Note: originally published on 01.04.2015; author Henry Overman

In their focus on delivering a good policy on time and on budget, people often do not think about evaluation until the end of a project (if at all). They then have to scramble around looking for 'evidence' to demonstrate its impact. Unfortunately, by this time it is usually too late to provide convincing evidence of impact because the design of the scheme does not allow for meaningful impact evaluation, or because basic data on the scheme (e.g. who got money) and how it operated (e.g. how decisions were made) are not available. In contrast, **good evaluation is embedded in the policy design process.** Sometimes this does happen in the UK (e.g. some of the high quality evaluations discussed in the NAO report on evaluation in government) but not for most local economic growth policies.

In fact, a large portion of the evidence base we have found on local economic growth policies comprises studies which look retrospectively at projects that have finished. Many of these studies are not even official evaluations – instead being undertaken by academics long after the policy has been implemented (and possibly completed). Evaluations after-the-fact can be useful – as our evidence reviews demonstrate – but there are many benefits to be gained from considering how a project will be evaluated from the earliest phases of project design. For example, deciding how the scheme would be evaluated means thinking about what data will provide the most accurate picture of the impact of the project. In turn, thinking about what data would capture success (or failure) focuses minds on which objectives are the highest priorities. This can help prevent the tendency to attempt to address too many diffuse goals with one policy – e.g., when thinking about evaluating an SME support programme we might think about collecting data on employment or firm productivity. Which is better would depend on whether the policy was trying first and foremost to improve profitability, or employment, or productivity. Asking what is the most robust method to evaluate progress against each of these objectives helps decide what is the main objective.

Thinking early about good evaluation approaches (such as RCTs or quasi-experimental approaches), can help design the programme and the evaluation in a way that provides more insight into how a policy works. In other words, this can allow us to evaluate what works better (i.e. how to improve effectiveness) rather than just focusing on what works (i.e. whether the programme had any effect). For example, in this Swiss employment support programme, vouchers for different monetary values were randomly assigned to businesses entering a programme. This design feature allowed the policy to be assessed not only in terms of overall impact, but also value for money. Policymakers were able to assess whether there were thresholds above which offering more money produced diminishing returns – knowledge which should allow for improved cost-effectiveness as the programme is developed. Importantly, everyone in this programme also received some assistance: clever design got round the common objection that randomisation leaves some people without help. Thinking about these issues carefully can also help improve policy processes – because it requires precision on exactly how decisions will be made about who gets what support.

Early clarity about how the project's impact will be measured also helps with the evaluation itself. Firstly, it allows for the use of more robust impact evaluation methods, while taking in to account other constraints on the policy design process. Second, it helps to ensure that the proper data is collected throughout the project life. Arrangements may need to be put in place to track participants over time, or count visitor numbers, or measure changes in profitability. This data will be more cheaply and accurately collected if good evaluation is built into the programme design, rather than if it has to be reconstructed later. It also allows data requirements for evaluation to be considered alongside those for monitoring – reducing duplication of time and effort if the two requirements are considered separately.

Finally, embedding evaluation in the design process forces us to think about how the results of the evaluation will feed back into future decisions about the programme. In the extreme, this might involve decisions about whether to continue funding the programme. Alternatively, evidence on what works better (e.g. the appropriate level of subsidy) can help us improve cost-effectiveness as a scheme is developed. Evaluations long after the programme is finished do not help with these two crucial decisions – and undermine the most important way in which evaluations may directly improve policy effectiveness for the specific policy being evaluated. Evidence reviews are useful, but certainly second best compared to a feedback loop direct from evaluation of a specific policy to the future development of that policy.

There are other advantages to embedding evaluation in the design process, but these are four of the big ones. The crucial message that emerges is to think about how and what you want to evaluate at the earliest stages of project development.

2. Define success

Note: originally published on 07.04.2015; author Henry Overman

The core objective of the What Works Centre for Local Economic Growth is to answer the question: which policies are most effective in supporting and increasing local economic growth? The answer, of course, depends on many things – not least what we actually mean by 'increasing local economic growth'.

In looking at the evidence, and talking to local decision makers, it is very clear that there is plenty of disagreement about what is meant by increased local economic growth. In our evidence reviews, we tend to focus on two specific outcomes: employment and productivity. That is, we use a fairly traditional (some would say narrow) definition of local economic growth. We also worry about whether effects come about because of improvements for existing firms and residents, or because of changes in the composition of the local economy.

Quite rightly, some people worry that employment and productivity are not the main objectives of some policies that we review – e.g. estate renewal or the provision of sports and cultural facilities. But the reason for focusing our reviews on these two outcomes is precisely because we want to help local decision makers understand the most cost-effective way to increase local employment or productivity. After all, that is our main remit.

All of that said, we also think it is very important to understand the wider effects of local policies on a whole range of outcomes. The crucial thing, especially when it comes to evaluating specific projects, is to carefully think through what effects we expect the project to have. To put it another way: **in order to determine whether or not a project is a success, we need to think about what success would look like, and therefore how it should be measured.** If, for example, estate renewal has objectives other than employment and productivity, then what are they and how would we know if we achieved them?

Unfortunately, even when the projects involved have a narrow economic focus, we consistently find that success is poorly defined. For example, in reviewing the evidence base for business advice effectiveness, we found a large number of assessments which were very muddled in stating the objective of a policy (if this was even defined). Policies often set out to support businesses to 'grow' – so far, so good – but when you start to measure business growth you have several choices. Are you interested in growing profits? Growing employment? Growing turnover? Output? Market share? Productivity?

Each of these factors is measured in a different way. More importantly, however, different forms of support will produce different kinds of 'growth'. The importance of clarity on objectives is not merely an evaluation concern. If we want business advice to serve the end of local economic growth, we may need to look beyond the effects at the business level. Increased profits or market reach may not have much impact at all on the local economy if they do not reflect increased productivity or translate into increased employment.

As we discussed previously in the 'starting early' section, one of the benefits of embedding evaluation into the design process is precisely that it forces us to think about how we might measure and define success. In turn, that forces us to think through exactly how we expect given policies to affect the local economy.

3. What to evaluate?

Note: originally published on 15.04.2015; author Henry Overman

The previous section with the issue of the importance of defining success when thinking about how to evaluate. It might seem that answering that question also answers the question of 'what to evaluate'. If the policy objective is employment, then we evaluate whether or not the policy has a positive impact on employment. If we are interested in cost-effectiveness, it is clearly helpful to undertake evaluations that address whether or not a policy works at all. It would be great if there were a large number of impact evaluations that clearly said that policy A works, policy B does not. Unfortunately, as our reviews to date make clear (e.g. on employment training or business advice) this is rarely going to be the case – some employment training programmes work, others do not (and similarly for business advice, access to finance, etc).

However, in practice, the challenge that we face in improving cost-effectiveness does not just come down to conflicting findings. It may sound cynical but even if the evaluation evidence on, say, business advice was uniformly negative, we would be amazed if policymakers suddenly stopped funding programmes to provide such advice. Partly this comes down to politics – cost-effectiveness is not the only consideration when it comes to policy decisions. More prosaically, individual policy makers are often given a budget to spend on a broad policy area with one or two objectives. For example: commission an employment training programme to reduce unemployment among young people.

The question that such a practitioner really wants answered is not 'what works' but 'what works better'? Given that we are going to provide employment training, what kind of training should we be providing? Short courses or long courses? On the job or off the job? Many considerations (resources, capacity constraints, etc) will have a bearing on how these questions get answered. In keeping with the Centre's objectives, we would argue that cost-effectiveness should play a central role in answering questions about these policy design features. While often not well understood, good impact evaluation should play a crucial role in answering such questions.

For example, when NICE provides guidance it does not try to answer the broad question about what makes us healthy. Instead, it tries to decide what treatments work best in addressing particular conditions. Similarly, the Educational Endowment Foundation focuses on assessing the effectiveness of very specific interventions (including after school programmes, arts participation, extended school time, feedback) on improving one specific outcome: the attainment of disadvantaged pupils of primary and secondary school age.

Trialling two or more versions of a policy is a very effective method of comparing their effectiveness. Some of these experiments can be very large scale. One recent academic paper describes a trial from France involving a total of nearly 44,000 unemployed individuals allocated in to three different groups. But trials can also be much smaller – which means that local policy experimentation can provide a great context in which to try to figure out what works better, particularly if a number of local areas are willing to collaborate in piloting different approaches.

We provide detail on how to structure a policy's implementation to provide comparison or control groups in the next section 'Find a control group', but first we would talk about defining the evaluation questions.

4. Find a control group

Note: originally published on 23.04.2015; author Henry Overman

Most policymakers are familiar with the basic approach to monitoring policy outputs (such as the number of people who have gone through a training programme, and how many got a job). However, in economic development we tend to be less practiced at isolating the causal impact of policy interventions.

By causal impact, the evaluation literature means the difference between the outcome for individuals 'treated' in a programme, and the outcome they might have experienced without it. Pinning down causality is a crucially important part of impact evaluation. After all, estimates of the benefits of a programme are of limited use to policymakers unless those benefits can be attributed, with a reasonable degree of certainty, to that programme.

Establishing causality requires the construction of a valid counterfactual – i.e. what would have happened to programme 'participants' (individuals, firms or areas) had they not been treated under the programme. The way in which this counterfactual is constructed is the key element of impact evaluation design.

A standard approach is to create a counterfactual group of similar individuals, firms or areas not participating in the programme being evaluated. Changes in outcomes can then be compared between the 'treatment group' (those affected by the policy) and the 'control group' (those not affected by the policy).

Another approach – useful if we are interested in figuring out 'what works better' – is to offer similar groups different treatments. For example, in the case of business support, we take two similar types of business and offer some mentoring support (more expensive) and others are referred to online materials (a less expensive type of advice). The two groups can then be tracked to see how the different approaches play out.

The fundamental idea here is to make sure we are comparing groups that are similar. Such comparisons help address concerns that other factors might be driving changes for participants in a programme. These other factors might be 'external' to the programme (for example, it is offered to struggling areas) or they might be 'internal' (e.g. only certain types of firms choose to take part in a business advice programme). There is a lot more detail which can be considered when choosing a control group – and our selection of case studies for particular policies provide examples of how evaluations have dealt with these issues in practice. But even for those with less technical expertise – thinking about suitable comparison groups is a crucial step in thinking how you might evaluate policy success, as well as what you can learn in the absence of a comparison.

If you find yourself convinced, but are struggling to convince partners, you may find Ben Goldacre's succinct description on Newsnight in February 2015 provides an invaluable introduction. And of course, we are always available to try to help those looking to undertake more robust evaluation.

5. Collect Data

Note: originally published on 29.04.2015; author Henry Overman

So far in this series, we have talked about the importance of starting early, of defining success, of thinking about what to evaluate and of finding a control group.

This section considers another key step – the importance of collecting data. It might seem simple, but the availability of appropriate data is often one of the biggest stumbling blocks, especially when trying to retrofit evaluation to programmes that are already well under way.

There are many points that could be discussed, but we highlight the three major ones.

First, you need information on the identity of programme participants. Which individuals, firms or areas have benefited from policy support? Many people would be surprised that such information is often not readily available – even for accountability purposes if nothing else. But we have seen numerous occasions where such information is not being collected (even when large sums of money are involved). It can be a particular problem when second parties hand out the money on government's behalf.

For what it is worth, such data should be gathered regardless of evaluation plans. But such data are critical to any attempt to understand the impact of policy. This information will need to be gathered and stored systematically and preferably in a way that allows it to be matched to other sources of information on participants (of which, more below). And if we truly want to get at questions of cost-effectiveness we also need to keep information on what kind of support participants have received (e.g. how much money) particularly if the kind of support can vary a lot across participants.

The second crucial data issue is to decide which data will best capture outcomes that are linked to the objectives of the programme. Thinking about how we define success should help clarify this. Once we have figured out the outcomes in which we are interested, there are two more things to think about:

- we need this data before and after the programme so we can see whether there has been any measurable change in the outcome for participants;
- we need similar before and after data for the control group that we are going to use as a comparison group for participants.

In fact, these two requirements – the use of before and after data, combined with a suitable control group – are key building blocks in our evidence reviews. Evaluations that do not have both of these fall short of the evidence standards we set for our evidence reviews as we discuss further in our scoring guide.

The third decision that we need to make is how we will collect this data on outcomes. There are essentially two possibilities. The first is to use bespoke survey data that is collected specifically for evaluation purposes. Unfortunately such data can be expensive. It also brings a temptation to try to collect large amounts of data – on the process, on a large range of outcomes, on whether people think the programme is making a difference, etc. This then leads to large sprawling evaluation reports which consider a huge range of issues. For most local decision makers it is hard to believe that the cost of bespoke data is justified.

We should focus attention on getting a smaller amount of data on key outcomes. At the end of the day, in combination with a suitable control group, information on only one or two outcomes such as changes in employment or wages allows us to answer the most important questions of all – did the policy work or is one kind of support more cost effective than another?

Another option for keeping costs down is to make much better use of secondary data. Using such data also helps address concerns about imposing a burden on the non-treated participants in the control group as the data is already being collected for other reasons (small bespoke surveys also help address that concern).

At the moment, using secondary data is easier for firms and areas than it is for individuals. For example, for firms, if we have information on firm names, company house identifiers and postcodes we can achieve very good matches with administrative data (such as the Inter-Departmental Business Register) which already provides information on key outcomes of interest (particularly employment). Even for individuals, efforts are being made to improve access to data sets – such as those held by the Department for Work and Pensions – which provide detail on a range of outcomes. Indeed, in a previous blog, Majeed Neky described how this data – combined with a randomised control trial – will hopefully allow a group of London councils to evaluate their 'Working Capital' active labour market programme.

Of course, one problem with these secondary data sources is that there is usually a time lag before they are available. Whether this is a major problem will depend on the time frame over which any effects of the policy will be felt. For many local economic growth programmes, where effects are expected to be longer term, this should not be such a problem – something we discuss more in the next section.

6. How long?

Note: originally published on 18.05.2015; author Henry Overman

In the previous section we considered some of the issues around collecting data. One related issue that we did not discuss is the question of when to evaluate?

In an ideal world, we would want to think about when policy effects were likely to occur – and evaluate accordingly. For example, if the aim of a training programme is simply to get people into work it might be appropriate to evaluate shortly after the programme was completed. But if we are interested in whether the programme leads to long-run employment, we might want to evaluate outcomes a year, or more, after programme completion.

In practice, unfortunately, there are often political pressures to provide early evidence on impact. It is easy to criticise this impatience and to call for politicians to take a longer term view. But it is important to try to move beyond this typical reaction – for two reasons. First, because experience suggests that appeals to take the long-term view will often fall on deaf ears. We do not want the case for better evaluation to stand or fall on an argument we cannot win. Second, if we change our thinking to see evaluation as something that is embedded in the policy design process then the politicians may be right to insist on early evidence of impact (even if they are doing it for the wrong reasons).

In medical trials, for example, it is good practice for the trial protocol to describe the procedure regarding decisions on discontinuation of the trial. With that parallel in mind, and given the political realities, it is important to think about ways in which we could get some short-term indication of programme effects (possibly from bespoke survey data, or from secondary data that provides imperfect indicators) that will help inform policy development while waiting for secondary data to become available.

To give an example, let's imagine that a LEP wants to provide information to Further Education students on the labour market outcomes they can expect if they take different types of courses. The idea being that telling them about, say, the high wages they will earn as an engineer might encourage them to take courses that lead to that career. The LEP decides to pursue a randomised control trial to evaluate this idea, before rolling out across all their colleges. To do this, they randomly send some new FE entrants information on likely labour market outcomes shortly before they choose their courses (there are issues with this very simple design, but we can ignore them for the purposes of this explanation). The long-run aim of this policy is to increase the number of people employed as engineers. But there is plenty of opportunity for using short run indicators to see whether these long run effects are likely to be realised. FE colleges will have information on course choices, on completion rates, on exam outcomes and on first destinations. Using this information allows indicative evaluation within a few weeks (course choices), a few months (completion/results) and around a year (first destination for a one year course). If we see no movement on any of these outcomes, we might want to avoid wasting further money on this project next year (as well as avoiding the costs of the bespoke survey planned to pick up career destinations two years on from graduation). Of course, if these indicators are moving in the right direction, that may well justify continuation of the pilot and the costs of following up with a longer time scale.

The crucial point is that we need to be thinking about how the evaluation will be used to feedback into policy design. If piloting is truly embedded in the policy design process then we may want to make compromises on suitability to allow early evaluation based on readily available data. Of course, the long-run follow-up that looks at the ultimate objective is still vital to assessing the long-run effectiveness of the policy. But looking at short-run indicators can provide important information that can help determine whether to keep going – a decision which has implications for both programme and evaluation costs. Once again, this discussion only serves to highlight the importance of starting early and embedding evaluation into the policy design process.

7. Plagiarise

Note: originally published on 29.05.2015; author Henry Overman

One of the major problems with improving impact evaluation and getting it embedded in the policy design process arises because people think that evaluation is too complicated.

Throughout this series on how to evaluate, we have tried to show that thinking about impact evaluation need not be a very technical exercise. Asking simple questions about what and when to evaluate helps to define success. Thinking about how we will measure changes for those on the programme and compare those changes to a control group allows us to understand whether we will be able to attribute success to the programme (as opposed to all the other factors that will be at play). While thinking through these issues from first principles is to be strongly encouraged, we should also recognise the importance of copying freely from the approaches adopted in existing studies. Indeed, in an ideal world, we would be able to draw on evidence from multiple randomised control trials before we even considered rolling out expensive interventions more widely.

Unfortunately, in practice, we are a long way from having that volume of high quality impact evaluation evidence available for any local economic growth policies. But our reviews do identify many evaluations that meet the Centre's minimum standards – and thus provide a possible template for developing evaluation strategies. For example, an RCT shows how we might evaluate a scheme that provides adult education vouchers with the aim of improving labour market outcomes. A second example describes an RCT which uses business advice to achieve a similar outcome.

Admittedly, things get trickier when we move away from RCTs towards more statistical approaches. Indeed, one of the nicest things about RCTs as an impact evaluation approach is that they are relatively simple to understand, even for the non-specialist. Indeed, the EEF even has a simple guide for teachers to implement basic RCTs in their own schools and classrooms! But even when we move away from RCTS there are lessons to be learned from existing impact evaluations – even if some of the statistical techniques are more tricky.

We also provide a methodology guide which allows organisations thinking about commissioning an impact evaluation to assess how different approaches might rank on the Maryland Scale that we use to rank studies for the purposes of our review.

In short – this is one area where plagiarism does not necessarily give the right answer, but it certainly helps getting there.

8. Get everyone on board

Note: originally published on 10.06.2015; author John Holden

The previous sections in this series look at how good impact evaluation should be designed into local economic growth policies from the start and how, even with a small amount of planning, this need not be technically difficult or expensive. This section takes a look at a part of impact evaluation that is all too often overlooked: the need to bring the people or organisations you expect to act on the results in along with the evaluation process. Many good evaluations have failed to have the impact they rightly deserve because they missed out this crucial step. Likewise, many poor evaluations have had disproportionate impact because they got this right. There are two key questions that this blog explores: who needs to be “on board” and when.

The answer to the second question is easier than the first: as soon as possible, ideally before the intervention has begun or the evaluation design finalised. The reason to bring people in early is that by far the easiest way to ensure buy-in to the evaluation process, and ensure the results are acted on, is by developing a sense of ownership from the start. A few days delay in finalising the evaluation approach can save months of wrangling over technical details when the results of an evaluation are approaching publication. It can also make the difference between an evaluation being acted on and resulting in real policy change, and it being left sat on the proverbial (digital) dusty shelf.

When engaging with policymakers early in the life of an impact evaluation, it is important to remember that most local economic growth policy making is short-term in the UK. Policymakers are unlikely to be interested in the design of an evaluation that is not going to report for perhaps two years or more as it does not meet an immediate need. However, once publication is approaching, and the evaluation has the potential to disrupt an agreed policy approach, interest will rise exponentially. It is best to pre-empt this and take the time to engage widely from the start, despite the indifference that may confront an enthusiastic researcher.

Deciding who to get on board is more complex. In local economic growth policy, the list of actors with a potential stake in an intervention is very broad. Rarely can one organisation act in isolation from its partners, as the interdependencies between different parts of the system are too great. The same is true of interventions and their evaluation. Engagement with four key groups of stakeholders is likely to be necessary:

- **The funding organisation:** When an intervention is externally funded it is likely that the funding organisation will require an evaluation and will in all likelihood fund it, so for purely pragmatic reasons it is helpful to have them on board with the approach. It is also the case that the funding organisation will usually be one of the key actors that the evaluation will be looking to influence, so their buy-in is essential.
- **The delivery organisation:** There is no question that good evaluation requires access to high quality data and in most cases the delivery organisation will collect this information. Getting buy-in from those responsible for collecting evaluation data can make the difference between data input being seen as an unimportant administrative burden and it being viewed as an essential part of project delivery and service improvement. Buy-in from the frontline can also provide invaluable soft intelligence to give more rounded evaluation findings. A business advice intervention may fail because of a fundamental policy design flaw or because the business advisor spent all their time on Facebook. Spending the time building trust with those delivering a service can identify what the real lessons learnt are.
- **Partner organisations:** In local economic growth it is highly likely that any impact evaluation is going to identify that not all the benefits and not all the costs of an intervention flowed to the organisation that funded it. More sustainable delivery models can be developed

if key partners are included in the evaluation process so that they can get an insiders understanding of the effect of an intervention on their organisations finances.

- **Key decision-makers:** In local economic growth, local authorities and local enterprise partnerships are likely to be important actors in deciding whether to scale up or scale down an intervention based on the evaluation findings. Early engagement can help ensure that they anticipate the evaluation findings and factor the timings of the results into their policy development cycles.

Finally, take some time to think about the person you will likely probably never meet but that is desperate to be on board with you: the researcher or policymaker searching for evidence of what works. By making your evaluation findings available freely online you can have a policy impact that goes way beyond the immediate local context.

9. Ask for help

Note: originally published on 02.06.2025; author Ana Tibocha-Nino

Evidence-based policy is at the heart of good practices in policymaking. Impact evaluation provides a key source of evidence to help understand what really works and we need more impact evaluation of local economic growth policies to improve policy development.

However, designing and carrying out impact evaluation can be intimidating for policymakers. Most economic development practitioners have never undertaken impact evaluation, and many will not have commissioned evaluations either.

Designing an impact evaluation can also be complex. The process involves many stages and there are different methodologies (some better than others), each with their own assumptions, requirements and limitations. Going through all the steps, choosing an appropriate methodology and finding a suitable comparison group can be difficult. These issues also make commissioning an impact evaluation and finding the right provider a challenge.

Fortunately designing an impact evaluation can be a collaborative process.

Start by consulting the evaluation, research or insights team in your organisation (including those working in other areas). They may be able to provide support and feedback on evaluation design, commissioning process, and implementation. Delivery teams, as well as other colleagues, can help define success, identify data requirements and availability, and provide ideas and feedback.

What Works Growth can also help. We offer evaluation support to central government and local government. We provide training, resources, and can often offer bespoke support. This can include early conversations about whether an intervention is suitable for impact evaluation. We can also offer advice and feedback on impact evaluation design, reviews of draft methods and more, all free of charge. For local government, we also coordinate an MCA evaluation network, where evaluation managers offer each other advice and share best practice. Other What Works Centres, peer organisations, professional networks, and universities may be able to share previous evaluation experiences and help think about what and how to evaluate.

In short, it's possible to ask for help and support at any step of the process.

what works centre for
local economic growth

This work is published by the What Works Centre for Local Economic Growth, which is funded by a grant from the Economic and Social Research Council, the Department for Business and Trade, the Ministry of Housing, Communities and Local Government, and the Department for Transport. The support of the Funders is acknowledged. The views expressed are those of What Works Growth and do not represent the views of the Funders.

June 2025
What Works Centre for Local Economic Growth

info@whatworksgrowth.org
@whatworksgrowth

www.whatworksgrowth.org

HM Government

UKRI Economic and Social Research Council