WW book chapter

The What Works Centre for Local Economic Growth: Some lessons from the first ten years

Danni Mason (What Works Centre for Local Economic Growth), Max Nathan (University College London) and Henry G. Overman (London School of Economics and What Works Centre for Local Economic Growth)

Disclaimer: The What Works Centre for Local Economic Growth is funded by a grant from the Economic and Social Research Council, the Department for Business, Energy and Industrial Strategy, the Department for Levelling Up, Housing and Communities, and the Department for Transport. The support of the Funders is acknowledged. The views expressed are those of the authors and do not represent the views of the Funders.

Introduction

What can policy do to increase local economic growth? This is a question that has challenged decision-makers and academics for decades. It's also a fundamental question today, when the UK government has made 'levelling-up' one of its priorities – but is also faced with slow growth and tight budgets.

Research and evaluation have a crucial role to play in providing answers and increasing the effectiveness of policy-making. Unfortunately, making sense and making use of the evidence is not easy, especially for those tasked with delivering better economic outcomes for their communities.

The What Works Centre for Local Economic Growth (WWG) was founded in 2013 to work with policy-makers and help address these challenges. This chapter summarises what we have learnt from our first three phases of activity.

Phase 1: Evidence reviews - what works?

An evidence centre needs evidence. So our first task was to review what is known about the overall impact of local economic policies. After developing, testing, and iterating our review methodology, we published 14 reviews in two years covering areas including employment training, business advice, estate renewal, broadband and area-based policies. We would argue that was good progress for a small team (roughly five full time staff) covering a vast policy literature. We made several key decisions early on that helped us do this, but which also had implications for our work that resonated well beyond this first phase.

1) Systematic Evidence Reviews

The first decision concerned the kind of reviews we would produce. For people to have some confidence in our objectivity we felt that the reviews had to be systematic – by which we meant rules-based and summarising all available literature that met those rules. But there are different ways to do systematic reviews. For example, the Alliance for Useful Evidence[1] distinguishes between 'exploratory' approaches – describing who's doing what in different places – and 'structured' approaches, which draw on studies which use more formal methods to test policy effectiveness.

[1] Gough D, Oliver S, Thomas J (2013) Learning from Research: Systematic Reviews for Informing Policy Decisions: A Quick Guide. A paper for the Alliance for Useful Evidence.

We decided to take a structured approach for several reasons. Most importantly, in 2013 exploratory work on local economic policy was common, with far fewer structured reviews – something which remains true today. Structured approaches also played to the comparative advantage of our core academic team. And structured approaches provided the best fit with the What Works Network requirements.[2]

2) A focus on impact evaluations

Our second big decision was to focus on a specific type of evidence – impact evaluation evidence. Impact evaluations seek to identify and understand the causal effect of policy interventions and to establish their cost-effectiveness. To put it another way, they ask: 'did the policy work', 'how did it work' and 'did it represent good value for money'?

While all the What Works Centres recognise the central importance of impact evaluation, some also draw heavily on other types of evidence. These include correlational evidence – which identifies factors which are *related* to the outcomes of interest, but does not demonstrate a causal effect – and process evaluation evidence, which is designed to investigate *how* interventions are delivered and explain *why* they are effective or not.

Both of these provide valuable complements to impact evaluation, but we felt that within local economic policy, the evidence gap was most pronounced for impact evaluation. We also felt it was important to have a clear focus on evidence which demonstrates whether a policy has been effective, which neither of these other evidence sources can do.

Traditionally, local economic policymaking has *not* drawn much on evidence which uses these methods. Governments around the world increasingly have strong systems to monitor policy inputs (such as programme spend) and outputs (such as counting programme participants). However, they are weaker on identifying policy outcomes (programme effects) and in particular, many government-sponsored evaluations that look at outcomes do not use credible strategies to assess the causal impact of policy interventions. By highlighting the importance of impact evaluations and demonstrating their uses, we hoped to achieve a change in culture, as well as shifting specific practices.

[2] https://www.gov.uk/guidance/what-works-network, accessed 2 February 2022.

Pinning down causality is a crucially important part of impact evaluation. To assess causal impact requires an estimate of the difference that can be expected between the average outcome for individuals 'treated' in a programme, and the average outcome they would have experienced without it. Establishing causality requires the construction of a valid counterfactual – i.e., what would have happened to programme participants had they not been treated under the programme. So, we organised our reviews around the credibility with which evaluations establish causality: our reviews include any evaluation that compares outcomes for people, firms or places receiving treatment (the treated group) after an intervention with outcomes in the treated group before the intervention, relative to a comparison group used to provide a counterfactual of what would have happened to these outcomes in the absence of treatment.

It has been argued that our evidence standards are set too high. For evaluations of some area-based initiatives – such as infrastructure and sports and culture programmes – finding suitable control groups is challenging. However, internal exercises we ran on evaluations for sporting and cultural events and facilities strongly suggest our evidence standards are not the problem.[3] Instead, a clear pattern emerged that most of the studies we found, particularly for events, modelled the possible impacts before they were in place ('ex-ante' appraisal). But precious few evaluations look at the actual impacts after the event ('ex-post' evaluation). A small but growing body of impact evaluations for transport - as covered by our transport evidence review - shows that there *are* impact evaluation methodologies which can be applied to interventions like these.

Arguably, the barrier is not that impact evaluation is not possible, but that it is seen as less necessary, or less appealing, in a policy area where there is particularly intense focus on up-front economic *appraisal* – the quantitative modelling of *potential* impacts. The focus on appraisal is not surprising: policymakers are, perhaps understandably, more interested in figuring out ways to assess new projects, rather than discovering whether their estimates were right after the money was spent. But for future policy to be credible, rather than just plausible, it needs to be built on an understanding of what has worked in the past.

3) <u>A focus on key economic outcomes</u>

[3] We found something similar when producing our toolkits. While these continue to focus on impact evaluations, they use a lower evidence standard. But this shift did not result in us having a much wider evidence base on which to draw. The key constraint is the lack of impact evaluation evidence, not the methodological cut-off which we use.

A third key decision was to emphasise outcomes that most directly capture change in a local economy: wages, employment, and productivity. There are, of course, many other important ways in which places differ. But ultimately, if policy isn't increasing wages, employment, or productivity it's hard to claim that it's improving local economic growth.

That said, we were still interested in other outcomes. Sometimes, because they act as proxies for the effect on local economic performance: for example, if business advice increases firm sales, then this may translate to improved local economic performance. But we always need to be careful with such proxies. For example, if business advice boosts domestic sales at supported firms, these sales may simply be at the expense of other firms nearby. Good for the supported firm, not so good for the unsupported. There are also policy areas – such as estate renewal and sports and culture – where our evidence reviews suggest that spending tends to have a limited impact on the local economy but which might have plenty of other potential benefits including environmental, health, crime, and wellbeing. And there may be indirect links from these benefits to our core outcomes (e.g., economic benefits from improved health).

4) Piloting, scoring, iterating, updating

Having made these big calls, and refined a list of topics with practitioners, we carried out two pilot reviews – one on employment training and the other on business advice – issuing calls for evidence and searching the academic literature, think tank reports and policy evaluations.

Even with our high evidence standards, the initial literature searches were vast: we covered evidence from all OECD countries, with no pre-set time limit (although we focus on more recent evidence when summarising findings). The first review, for example, generated a longlist of over a thousand studies. We narrowed this down to a shortlist by sifting on titles and abstracts, then conducted a full appraisal of each shortlisted study (a few hundred), collecting key results and giving robustness scores – based on an adjusted Maryland Scientific Methods Scale – that reflected both the quality of methods chosen and quality of implementation.[4] We also developed detailed scoring guidance, in consultation with our academic panel[5] and tested it using double-blind trialling of scoring by team members. It's important to note that the ranking of individual studies is not an exact

[4] https://whatworksgrowth.org/resources/scoring-guide

[5] We set up a rotating panel of academic experts at LSE and other UK institutions, with specific experience in one or more policy areas and in impact evaluation of those policies.

science and often involves a degree of judgement. Indeed, anyone who has attended an academic seminar will know the extent to which such issues can be hotly disputed. Nevertheless, on average our scoring tended to produce rankings on which many evaluation experts would broadly agree.

Crucially, we allowed the way we classified and presented findings to evolve, as we looked for the most helpful approach to present evidence and lessons across different subject areas.[6] When we published what turned out to be our final Evidence Review, we revisited the early Reviews and brought our findings into line with our latest methodology. Fortunately, in practice, this mostly resulted in minor changes to our findings. In several cases these changes strengthened our initial findings, although they also weakened others due to a change which meant we classified more studies as having mixed findings rather than positive or negative.

Phase 2: Toolkits - what works best?

In 2016 we entered a second phase. Our evidence reviews provided a solid base for understanding the impact of a variety of individual local economic policies including employment training, business advice, estate renewal, broadband and area-based policies. Overall, they suggested that deep scepticism about policy effectiveness is not justified: in just about every evidence review, around half of the high-quality evaluations that we reviewed find positive effects on local economic outcomes, the remainder mostly find zero effects with a few finding negative effects.

At this point we made our first major pivot, away from systematic reviews towards more specific 'toolkits'. That is, we moved from answering the high-level question – 'does this policy work?' – to more detailed questions of policy design – 'in this policy area, what works best?'

There were two reasons for this shift. First, the fact that the evidence base was mixed for many of our topics – with some of the studies showing positive results and others not – meant that the answer to the first question could be uninformative, while there was much that could be usefully said in response to the second.

Second, at a practical level, many of our users were responsible for spending programme budgets. Rather than deciding whether to spend money on a given policy area (as users in central government

[6] We also reluctantly decided against trying to summarise value for money, given the lack of evidence.

sometimes do), they wanted to know *how best* to spend the resources they *did* have for a particular policy area that had already been chosen.

In many cases, the reviews found enough evidence to give additional guidance of this type. For example, our employment training review suggests that in-firm, on-the-job programmes outperformed classroom-based ones, and short training courses may be more cost-effective than longer ones; our access to finance review found some evidence that loan finance programmes were more likely to be successful than equity finance programmes; and our review of Enterprise Zones suggested that local employment conditions could influence the extent to which employment effects were felt locally. Our toolkits – which constituted the bulk of our work in this phase – are designed to summarise this type of evidence where it is available and make it accessible to users with no background in evidence.

But the toolkits are far from comprehensive. Anyone looking to develop an effective employment training programme will want to know about the costs and benefits of short and long-term approaches, but they may also want to know, for example, about the selection of participants, about how they are assigned to a specific training programme, and about how they are supported during their training. We do not have high-quality evidence to answer all of these questions.

Of course, appetite remains for simpler answers to the big questions: Does transport investment work? Do regeneration schemes work? Do cultural interventions work? Questions this broad are hard for all What Works Centres to answer, but the limited availability of high-quality impact evaluation for some types of local economic policy makes it a particular challenge for WWG. Compared to education, for example, the question of whether education interventions 'work' may seem as difficult to answer as the question of whether transport investment 'works'. However, for education, it is possible to fund many smaller trials, to dig down into the evidence, to examine what might explain the variation in results across studies and to focus on particular outcomes. You can't say whether education interventions 'work' in general, but you can say that education interventions which teach 'phonics' *do* work to improve literacy, based on data from hundreds of studies. Unfortunately, for many policies, the local growth evidence base is not developed enough for us to delve down and provide these more nuanced answers. We also think they undermine any suggestion that the mixed nature of the evidence base means that 'nothing works' when it comes to local growth.

Phase 3: Demonstrators and evaluation support - helping to fill the evidence gaps

Our work in Phases 1 and 2 delivered 14 reviews and over 30 toolkits. But many gaps remain. There are several challenges in *generating* impact evaluation evidence, both structural and practical, and many of these are particularly pronounced for local economic policy. We flag a number of these, then discuss the WWG's evolving response.

The 'gold standard' of impact evaluation is the randomised controlled trial (RCT). Fundamentally, randomisation is challenging for capital investments (like transport infrastructure) and for area-based economic initiatives (such as enterprise zones). For these, the size, lumpiness and longevity of interventions makes experimentation tricky (to say the least). However, this does not necessarily rule out high-quality evaluation. Quasi-experimental methods currently available (and more being developed) exploit pre-existing features of the policy, such as an eligibility cut-off, or staged roll-out to construct counterfactuals. More traditional approaches, such as propensity score matching, use data on non-participants to the same end. [7] Such approaches are often applied even when randomisation would be feasible but has not been implemented.

All these approaches to impact evaluation are technically complex, and not usually straightforward for local government to implement. Even *commissioning* high-quality evaluation requires some evaluation expertise – that's why central government employs many analysts and researchers to do just that. It's very hard to buy-in this type of expertise on a temporary basis, and many local partners find it impossible to justify permanent staff given financial constraints. Those larger places which do have this type of capacity are better placed to deliver impact evaluations that can help to fill the evidence gaps.

Another practical issue is that impact evaluation almost always has resource implications.[8] These include not just the financial cost of commissioning or delivering an evaluation, but also the additional resource required on the delivery side when evaluation has design or data collection

[7] As with other policy areas, we also need to think about how changes to the policy cycle could help improve decision making for capital projects. Rather than the experimentation that we highlight for other 'revenue' policy areas, we argue the focus for capital investment such as transport projects needs to be on better linking appraisal and evaluation. Considerable resources are devoted to the ex-ante cost-benefit appraisal of transport schemes as part of the decision-making process. The "ex-ante" appraisal framework should be complemented by strengthening guidance on the "ex-post" evaluation of transport investment.

[8] Although organisations like BIT have demonstrated that impact evaluation need not always be prohibitively expensive or time-consuming.

implications. Relatedly, local delivery partners can also be resistant to impact evaluation if they feel that the resources required for evaluation come at the cost of delivery. Both these issues affect the local economic policy especially hard, particularly given the impact of significant funding cuts to councils since 2010.[9]

Central government has a role to play in solving these issues, for example by coordinating impact evaluations across local areas or by providing sufficient resources as part of central government funding schemes, earmarked for evaluation at the outset. Coordination between different places doing similar interventions may also help solve a frequent methodological constraint faced by local evaluations when the size of schemes is small.[10]

There are also political and presentational challenges. The downside risk of being found to spend public money on something that 'didn't work' is substantial, while an upside to positive findings is often missing or limited. While this applies to many other policy areas, the balance of costs and benefits may make local evaluations particularly tricky. For example, centrally funded interventions which are shown to have been effective come with no guarantee of continued central support – what's the point of knowing something works if you can't afford to carry on doing it?[11]

Randomisation presents a particular challenge as it is often viewed as unfair or politically unfeasible to give some eligible individuals, business, or communities an intervention and not others, even when the intervention is not known to be effective. There *are* cases where central government has built in random allocation to the policy, but we know of no examples for local growth programmes led by local government.

A simple illustration of the fact that there are significant barriers and disincentives to doing high-quality impact evaluation of local growth interventions at the local level comes from a recent mapping exercise. Between December 2019 and March 2020, we contacted local authorities, combined authorities, local enterprise partnerships and central government departments, as well as other organisations all over the country to see what evaluations they had undertaken. In total, we sent emails to 1,329 people from 416 different organisations, and also conducted in-depth interviews with some of the respondents. A call for evidence, including to all local and combined

[9] https://www.instituteforgovernment.org.uk/explainers/local-government-funding-england

[10] When sample sizes are small it can be difficult to separate treatment effects from other things impacting the treatment and control groups.

[11] More conceptually, but importantly, local evaluation produces a public good because other areas can learn from evaluation. When benefits spillover, but the costs of evaluation are borne locally, then this public good will be under-provided.

authorities and local enterprise partnerships, resulted in 46 evaluations from 31 different organisations. None of these met the minimum standards for robust evaluation used by WWG.

The WWG has gradually evolved its 'offer' to users in the face of these challenges, complementing training and capacity-building with dedicated support on demonstrator projects, and in Phase 3, a dedicated fund for delivering evaluations. There were two main reasons for this pivot. First, in many policy areas we had gotten about as far as we could with the existing evidence base (although we have still added new areas). Second, because we discovered – through hard experience – that training and materials for evidence generation were not enough to get new evaluations off the ground in serious numbers without direct support and funding.

All WW centres face evidence gaps, but only a few are resourced to carry out the new evaluations needed to fill those gaps. The Education Endowment Foundation for example, is spending down a £125 million endowment on hundreds of evaluations of specific education interventions designed to improve pupil attainment. EEF identify evidence gaps (for example, what works to ensure pupils don't fall behind during the transition to secondary school) and issue public calls for relevant interventions which they can evaluate. They then fund schools to deliver a range of promising interventions as randomised control trials, so that high-quality evidence of their effectiveness can be collected and added to the evidence base.

Unfortunately, such an extensive exercise is beyond the budget of most of the What Works Centres. For these other centres, their primary option for filling evidence gaps is to identify promising interventions at the design stage – or, failing that, interventions already being delivered – and encourage the funders or deliverers to commission a high-quality evaluation. We support this approach by providing pro-bono support via our Evaluation Panel, composed of experts covering a range of policy areas. Supported by our funders, we have also created a 'stimulus fund', which allows us, for the first time, to cover the financial costs of a small number of evaluations undertaken with local partners, rather than simply providing support and advice.

In parallel with enhancing our offer to local partners, we have increasingly moved to working directly with central government programme funders. This second, more gradual pivot has also been essential in evidence generation because it has allowed us to influence the monitoring and evaluation frameworks around economic development programmes such as the Levelling Up Fund.

We think there is still more that could be done to leverage this existing spending on local growth interventions to get more good evaluation. For example, the current batch of central funds aiming to deliver levelling-up could be designed to support impact evaluation of the vast array of local growth interventions that they will fund. This can be done partly by building better incentives (and requirements) for evaluation into the design of the funds. In some circumstances it also requires central government to play a coordination role that no individual local partner can play.

Conclusion: Where are we now?

A lot has changed in the last 10 years. When we started, the evidence base was much thinner than most people would assume in some areas. For example, when we published our transport evidence review, in 2015, we found just 11 studies that looked at the economic impact of rail investments and met our evidence standards. As we write in 2022, our recent rapid evidence review on rail investment found an additional 18 studies. Our business advice evidence review found no UK based RCTs of business support. By 2019 the first two rounds of the BEIS Business Basics Programme had funded 11 trials and 26 proofs of concept. In short, more impact evaluations are happening.

But the barriers to local evaluation remain substantial. Our offer of technical advice through the evaluation panel and even financial support for demonstrators has not yet delivered a step change in the number of local growth evaluations of the type that we have seen for education evaluations over the last ten years. We need to undertake far more high-quality evaluations of the large amounts that we spend on local growth. We also need to place far more emphasis on experimentation, piloting, and evaluation at the early stages of policy development and on ensuring that lessons learned feed back into decisions about scaling up (or down) and into improving policy design. We suspect that additional resources and coordination and improved incentives will all be needed to achieve such a shift.

While there's still a long way to go in shifting people's understanding of the role that evaluation can play in helping improve local economic policies, a small but growing group of practitioners in both central and local government are trying to improve evaluation and embed it in the policy design process. Pre-pandemic, we put in place our evaluation panel and stimulus fund to respond to this growing demand. Hopefully, our attention can soon turn back to supporting efforts to better evaluate and improve our understanding of what works for local economic growth.