How can we improve the evaluation of (urban and regional) policy?

Prof Henry G. Overman (LSE, WWC)
Introduction
Overview

• What do we know about policy effectiveness?
  – Systematic reviews from the What Works Centre for Local Economic Growth

• Higher quality evaluation

• Improving the evaluation of policy
Systematic reviews: What Works Centre for Local Economic Growth
Our aim

Significantly improve the use of evidence in the design and delivery of policies for local economic growth and employment – leading to more effective policies and policymaking.

Our main audience are local government, and parts of central government that interact with them.
What Works Centre

- Evidence reviews
  - Capacity building
  - Demonstration projects
  - Understanding what works
    - Capacity building
    - More effective policy
Impact evaluation

Evaluating impact

Change in outcome for those in the programme vs Change in outcome for those not in the programme
To identify what works, each evidence review sifts and assesses the evidence to find evaluations which are robust and clearly identify policy impact. We do this using a 5 stage process.

1. **Scope**: Academic panel

2. **Search**: Existing literature and evidence is reviewed on the basis of an agreed review question, specific search terms, and a set of inclusion criteria

3. **Sift**: Evaluation evidence is collected using a wide range of sources

4. **Score**: The full set of evidence is refined based on its relevance and the robustness of the research method

5. **Synthesis**: Each study is scored based on the quality of method and quality of implementation

Conclusions drawn are based on a combination of these findings and existing literature
<table>
<thead>
<tr>
<th>Policy</th>
<th># Studies</th>
<th>SMS3</th>
<th>Emp.</th>
<th>Positive</th>
</tr>
</thead>
<tbody>
<tr>
<td>Access to Finance</td>
<td>1450</td>
<td>27</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Apprenticeships</td>
<td>1250</td>
<td>27</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>Broadband</td>
<td>1000</td>
<td>16</td>
<td>10</td>
<td>5</td>
</tr>
<tr>
<td>Business Advice</td>
<td>700</td>
<td>23</td>
<td>17</td>
<td>8</td>
</tr>
<tr>
<td>Employment training</td>
<td>1000</td>
<td>71</td>
<td>65</td>
<td>33</td>
</tr>
<tr>
<td>Estate renewal</td>
<td>1050</td>
<td>21</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Innovation</td>
<td>1700</td>
<td>63</td>
<td>10</td>
<td>6</td>
</tr>
<tr>
<td>Public realm</td>
<td>1140</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Sports and culture</td>
<td>550</td>
<td>36</td>
<td>16</td>
<td>4</td>
</tr>
<tr>
<td>Transport</td>
<td>2300</td>
<td>29</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>EZ/EmpZ</td>
<td>1300</td>
<td>30</td>
<td>27</td>
<td>15</td>
</tr>
<tr>
<td>EU SF</td>
<td>1300</td>
<td>18</td>
<td>11 (GDP)</td>
<td>5</td>
</tr>
</tbody>
</table>
Limits to evidence

• Small numbers of studies reaching SMS3 threshold
• Low percentage of those that look at outcomes of direct interest (e.g. employment, productivity)
• Half (or less) find evidence of positive impact
• Little evidence on what aspects of policy design make a difference

• Half the schemes that are properly evaluated have no measurable impact on policy objective
• Many schemes are not properly evaluated
Detail: e.g. broadband

• Broadband can affect firm productivity, number firms, and local labour markets (e.g. employment, wages)

• Effects not always positive, not always large, and may depend on complementary investments by firms (e.g. training, re-organising supply chain)

• Effects vary across different types of industries workers, areas (svc > manuf; skilled > low skilled; urban>rural).
Detail: e.g. transport

• Road projects
  – can positively impact local employment. But effects are not always positive & majority of evaluations show no/missed effects
  – may increase firm entry (either through new firms starting up, or existing firms relocating). Doesn’t always increase overall number of firms (new arrivals may displace existing firms)
  – tend to have a positive effect on property prices, although effects depend on distance to the project (and the effects can also vary over time).
## Toolkits: employment training

<table>
<thead>
<tr>
<th>Policy Design Tool</th>
<th>What does it aim to do?</th>
<th>How secure is the evidence?</th>
<th>How much does it cost?</th>
<th>How effective is it?</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Careers Counselling</strong> helps individuals choose appropriate training</td>
<td>Improve take-up and completion of training</td>
<td></td>
<td>£££</td>
<td>⭐⭐⭐⭐</td>
</tr>
<tr>
<td><strong>Financial incentives</strong> are payments offered before, during or after training</td>
<td>Improve take-up and completion of training</td>
<td></td>
<td>£££</td>
<td>⭐⭐⭐⭐</td>
</tr>
<tr>
<td><strong>Pre-qualification</strong> courses are a pathway to further education or training</td>
<td>Improve performance and completion of training</td>
<td></td>
<td>£££</td>
<td>⭐⭐⭐⭐</td>
</tr>
<tr>
<td><strong>Reminders</strong> provide people with information about their training by text or email</td>
<td>Improve attendance to training programmes</td>
<td></td>
<td>£££</td>
<td>⭐⭐⭐⭐</td>
</tr>
</tbody>
</table>
## Toolkits: business advice

<table>
<thead>
<tr>
<th>Policy Design Tool</th>
<th>What does it aim to do?</th>
<th>How secure is the evidence?</th>
<th>How much does it cost?</th>
<th>How effective is it?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mentors</td>
<td>are experienced business professionals who provide advice to SMEs</td>
<td><img src="#" alt="Security" /> <img src="#" alt="Security" /></td>
<td>£££</td>
<td><img src="#" alt="Effectiveness" /> <img src="#" alt="Effectiveness" /> <img src="#" alt="Effectiveness" /> <img src="#" alt="Effectiveness" /></td>
</tr>
<tr>
<td>Public advisors</td>
<td>offer advice on starting up or running a business</td>
<td><img src="#" alt="Security" /> <img src="#" alt="Security" /></td>
<td>£££</td>
<td><img src="#" alt="Effectiveness" /> <img src="#" alt="Effectiveness" /> <img src="#" alt="Effectiveness" /> <img src="#" alt="Effectiveness" /></td>
</tr>
<tr>
<td>Subsidised consultancy programmes</td>
<td>provide grants or vouchers for firms or entrepreneurs that cover all or part of the costs of private consultancy services</td>
<td>Improve firm performance and increase number of start-ups</td>
<td>£££</td>
<td><img src="#" alt="Effectiveness" /> <img src="#" alt="Effectiveness" /> <img src="#" alt="Effectiveness" /> <img src="#" alt="Effectiveness" /></td>
</tr>
<tr>
<td>Tailored support</td>
<td>provides advice tailored to the requirements of a business</td>
<td><img src="#" alt="Security" /> <img src="#" alt="Security" /></td>
<td>£££</td>
<td><img src="#" alt="Effectiveness" /> <img src="#" alt="Effectiveness" /> <img src="#" alt="Effectiveness" /> <img src="#" alt="Effectiveness" /></td>
</tr>
<tr>
<td>Training</td>
<td>involves publicly funded courses for existing firms or for individuals aiming to start a business</td>
<td><img src="#" alt="Security" /> <img src="#" alt="Security" /></td>
<td>£££</td>
<td><img src="#" alt="Effectiveness" /> <img src="#" alt="Effectiveness" /> <img src="#" alt="Effectiveness" /> <img src="#" alt="Effectiveness" /></td>
</tr>
</tbody>
</table>
Higher Quality Evaluation
The problem: many ‘Evaluations’ as exercises in self-justification

- Count ‘jobs created’, ‘businesses created’, etc
- But many – even most – may be diverted from nearby areas (Displacement) e.g. Enterprise Zones
- Many might have been created anyway (Deadweight) e.g. early intervention for newly unemployed
- May be added jobs through local multiplier
- And – given net job creation – cost per job?
Policy Evaluation

At some level applied common sense

1. Need to define objectives
2. ‘Measure’ impact of intervention on outcomes
3. Value costs
4. Value outcomes
5. Recognise that policy is an investment
6. Remember there is an opportunity cost to funds
The counterfactual

• Worthwhile evaluation studies must devise ways to estimate what would have happened in absence of policy.

• The most important element in judging how credible is an evaluation study:
  – how credible is method used to provide the counterfactual.

• Job change, unemployment, growth result from the combined effect of many factors – so need to isolate effect of policy from other influences.
Participants vs non-participants

• Why not just compare outcome for participants with outcome for non-participants?

• Selection problems
  – Individuals self-select into programmes
  – Admin may select on basis of characteristics
  – Dropout
  – Non-random programme placement

• This is a problem if selection is on the basis of characteristics that affect outcomes
  – e.g. Most able unemployed get training

➤ Crucial problem: finding valid control group
Improving Evaluation
Rigorous evaluation is possible for spatial policy

• Randomisation
  – Portas Pilots (should have been …)
  – Working capital programme
  – Mental Health Trailblazers
  – Growth hubs?

• Look for things correlated with participation but uncorrelated with outcome e.g. rule change (IV)
  – e.g. Redrawing of selective assistance map for RSA
Rigorous evaluation is possible for spatial policy

- Look for sharp breaks on who gets treated (discontinuities)
  - Boundaries of programme areas (compare either side) e.g. LEGI
  - Timing (compare early interventions to later) e.g. SRB commercial buildings
  - Size limits (compare size x+1 to x-1) e.g. French business support
Rigorous evaluation *is* possible for spatial policy

- Selection on unobservables
  - Before and after for participant and non-participant (difference-in-difference)
  - Urban renewal programmes in Berlin

- Selection on observables
  - Regressions, propensity score matching, etc
  - Earlier evaluations of RSA
Rigorous evaluation *is* possible for spatial policy *but* …

- Requires evaluation to be embedded in design
- Very hard to reverse engineer once policy already in place (which explains why so many bad evaluations out there)
The policy cycle

Monitor

Implement

Evaluation

Assess and approve

Design
Embed evaluation in design

1. How monitor
2. How implement
3. How evaluate
4. Improve
5. Test?
Rigorous evaluation *is* possible for spatial policy *but* …

- Needs short term win for decision makers but also allow for medium to long time horizon
- Needn’t be costly if good monitoring data and policy details made freely available (‘open data and evaluation’)

Rigorous evaluation won’t be possible for all of spatial policies

• Limitations (e.g. for UK Growth Deals)
  – Estimate of aggregate, national Growth Deals impact
  – Estimate of overall Growth Deal impact at local level
  – Single method for assessing all aspects of Deal
  – Complete coverage of all activities

• Trade-off monitoring and evaluation
  – Robust monitoring plus evaluation of specific activities
  – Clear understanding of how these activities contribute to outcomes/objectives
When rigorous evaluation *isn’t* possible

- Be very wary about relying on low quality techniques (e.g. self-reporting)
- May be better to think about effective (and cheaper) monitoring
Conclusions

• Robust evaluation of (some) components of local economic growth policies is possible
• Focus on improving evaluation for some aspects of policy
• Importance of embedding evaluation (and testing) in the design process