Revisiting Brussels Workshop

Elliot Stern

Helpdesk Evaluation Training
Lake Balaton Hungary 23rd May 2018

Evaluating Results and Impacts

- My aim is not to make you into an IE expert but to enable you to have sensible conversations with evaluation contractors, colleagues and other experts
  - but we cannot avoid technical content entirely….

Evaluating Results and Impacts

This includes

- Drawing up terms of reference

- Assessing proposals that cross your desks

- Managing and steer ongoing evaluations

- Assessing the strength of conclusions and recommendations reached by those conducting evaluations

- Developing new programmes and policies that are 'evidence-based', i.e., learn lessons from completed evaluations

An emphasis on results & impacts

- The current programming period sets out to strengthen the result-focus of EU programming (Article 56(3) CPR)

- This is associated with encouragement for 'impact evaluations' that set out to examine *whether*, the *extent to which* and *how* a programme caused the intended effects.

Evaluating Results and Impacts

Different definitions of 'results'

- Policy makers – often visionary and over-optimistic: we will transform our economy

- Public authorities – bureaucratic and administrative: the programme met its targets and produced credible indicators

- Stakeholders – how to weigh different interests: you say it was a success but what about young people

- Beneficiaries – short term and self-interested: has my enterprise benefited

- Socio-economic partners – longer-term and generic: have new value-chains been created

Evaluating Results and Impacts

- 'Impact Evaluation' (IE) is often associated with a particular methodology: the pioneers of IE who often came from statistical and experimental research traditions – so for them IE= experiments/RCTs
- We do owe the counterfactual school of evaluators a great deal: they made us think about what produces programmes results

Can we demonstrate that the programme caused the intended outcomes?

Evaluating Results and Impacts

- This concern with causal attribution - being confident that results really are produced by programme inputs, not by other things is important – so is explaining 'why' & 'how' results occur

- For example: the economy improved and unemployment would have gone down anyhow; the grant went to already successful SMEs; increases in innovation were the result of previous investments in R&D; city-centre renewal was caused by a new retail park

- IE sets a much higher standard of 'proof' that results were really caused by the programme

Defining Impact Evaluations

Impact Evaluation (IE) in policy settings sets out to do three things:

- First to demonstrate that a programme *'caused' an 'effect'* – the intended results

- Second IEs are often expected to explain *how* a programme works

- Third to consider the *contribution* a programme makes

Evaluating Results and Impacts

- Impact evaluation is now firmly about causal attribution BUT today we acknowledge that there are various methodologies to assess the relationship between cause and effect

- We have experimental methods, *and* what are called 'Theory-Based Evaluation' now also recognised in Commission guidance

- There are also other new and emergent evaluation approaches – statistical modelling; comparative case-studies; simulations; network analysis; agent based modelling...

- But there is a shortage of available skills using new methods

Evaluating Results and Impacts

Although we are interested in results, impacts, linking cause and effect we do this for a purpose – not simply to make causal claims!

The main purposes of evaluation remain as they always are:

- Accounting for Results
- Better Management and Delivery
- Learning and improvement
- Capacity development and sustainability

In order to improve programmes only looking at results is not enough

Designing Evaluations

- A 'design' is more than a method

- Interview surveys, case-studies, statistical analyses of labour market data, observational studies, all can be part of different designs

- A design involves a deep understanding of what we want to know, the programme context in which we are working and of the capabilities of different families of methods

- On the basis of this understanding we can consciously choose which of the many ways we could evaluate any programme

The Design Triangle
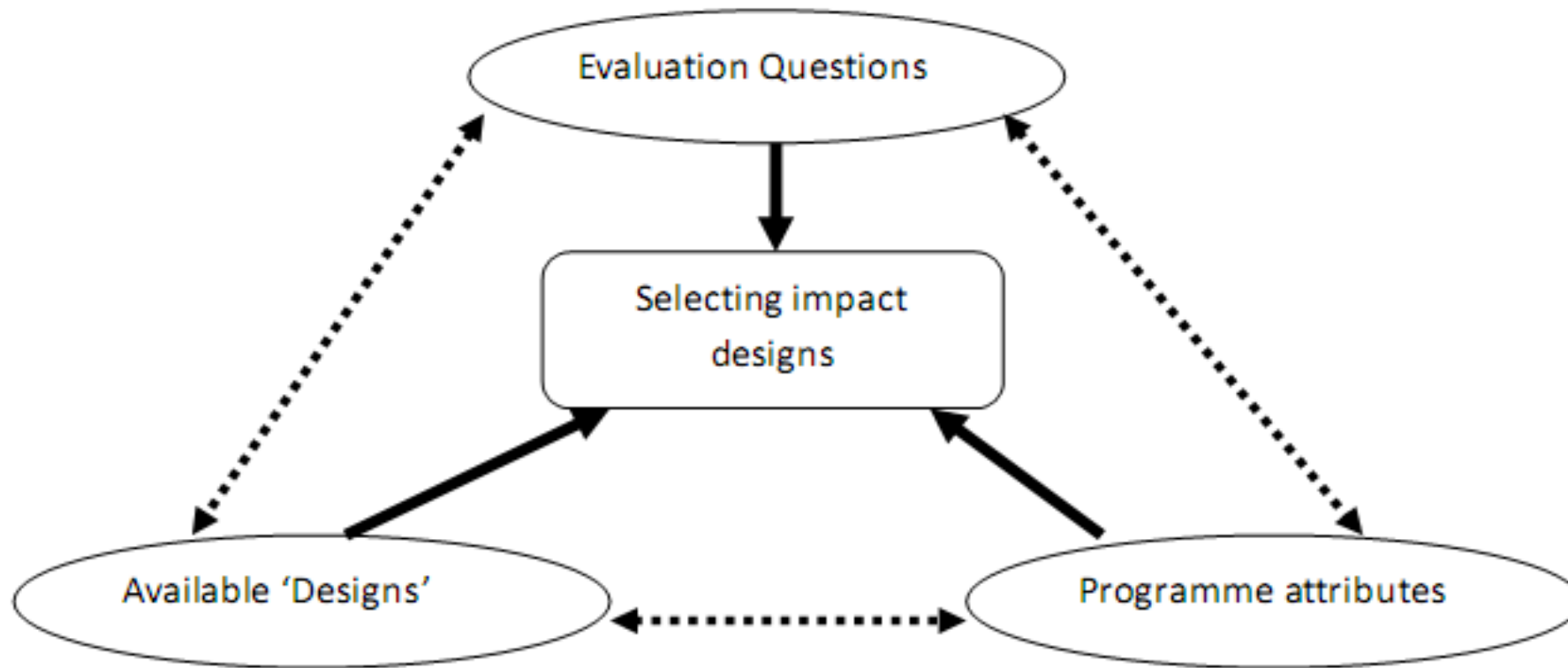
Designing Evaluations

The 'Design Triangle' suggests we need to match:

- Evaluation Questions (what we want to know)

with the

- Characteristics of Programmes (the 'object' to be evaluated)

with

- Available designs and their capabilities (what these designs can do given the Evaluation Questions being asked and the characteristics of Programmes)

Importance of Evaluation Questions

- EQs are not the same as an 'interview question' – they are 'high-level' question able to be analysed and studied

- EQs focus on the relationship between a programme and real world 'consequences' 'effects' and 'results'

- EQs that only focus on the internal operations of a programme and do not try to illuminate the relationships between programmes and real-world effects are not EQs!

- Most evaluations will have a number of prioritised EQs – each addressing a specific point of interest in a programme

Programme Characteristics

- Evaluation Questions have to take account of programme characteristics

- There are many ways to characterise a programme – in terms of sectors; objectives, their degree of innovativeness etc.

  - For the purpose of specifying EQs, characteristics should first be understood in terms of an 'intervention logic', 'theory of change' or 'programme theory' set into a wider context

  - Theories of Change can be variously described but usually cover the sequencing of a programmes cycle set into a wider socio-economic and regional context

Programme Characteristics
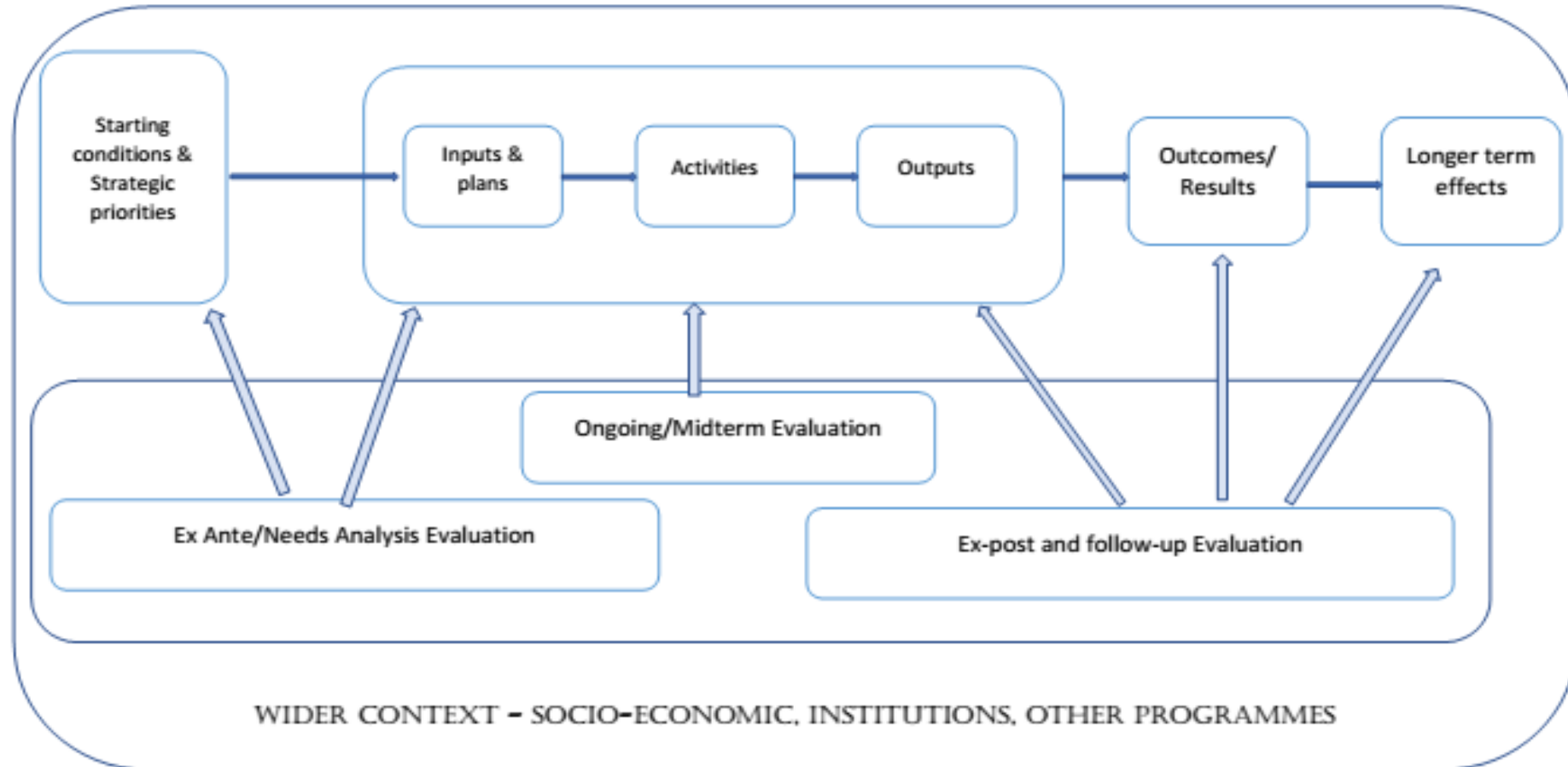
Programme Characteristics

When we discuss methodological choices there are other kinds of programme characteristics that also have to be considered. These include for example:

- Innovativeness of programme goals and delivery
- Simplicity or complexity of the intervention
- Types of outputs envisaged – whether they are material, behavioural, new services, new institutional arrangements
- Timescales and trajectories of change
- How bounded or embeddedness programmes are in relation to other programmes, activities and systems

Accounting for Results

At a *high level* we may want to know:

- Have programmes achieved their objectives?

In greater detail we may want to know:

- Can results be shown to be consistent with the ToC/Intervention Logic?
- Can we demonstrate that the programme caused the result?
- Who were the programme beneficiaries & how did they assess its results?
- Have results occurred consistently and if not, how can we explain differences across programmes?

Four Types of Causal and Explanatory Designs

- **Regularity frameworks** that depend on the frequency of association between cause and effect - the inference basis for statistical approaches to IE

- **Counterfactual frameworks** that depend on the difference between two otherwise identical cases – the inference basis for experimental and quasi experimental approaches to IE

- **Multiple causation** that depends on combinations of causes that lead to an effect - the inference basis for 'configurational' approaches to IE

- **Generative causation** that depends on identifying the 'mechanisms' that explain effects – the inference basis for 'theory based' and 'realist' approaches to IE

Strengths and weaknesses of different designs

'*Regularity*' requires high numbers of diverse cases to capture sufficient diversity (or difference) and numbers of variables for measurement

Counterfactuals are good at answering the question: 'Has this particular intervention made a difference here?' But weak on external validity questions: 'Will it work elsewhere?'

Generative causation is strong on explanation but weak on estimating quantities or extent of impact.

Experiments and *regularity/statistical association* approaches work best when causal factors are independent, but not if causal factors interact

Neither experiments nor statistical models are good at taking account of cultural, institutional, historical and economic settings but are good at measurement in the right circumstances…..

Returning to Evaluation Questions –
Not all IE questions are the same

1. To what extent can a measurable impact be attributed to this intervention?

Experiments and statistical models

2. Did the intervention make a difference or contribute?

Process tracing and Contribution Analysis

3. How has the intervention made a difference?

Theory-based and structured case-based methods

4. Will the intervention work elsewhere/elsewhen?

Methods that take context seriously – Contribution analysis & Realist approaches

Limitations of the Single Evaluation

- Most of the designs and methods discussed so far have related to single cases

- Most Programmes include multiple cases – firms, communities, infrastructures and service providers; and every Member State have multiple programmes each containing multiple cases

- We know that learning and reliable conclusions through evaluation, also has to be built on multiple cases – what we call synthesis methods

- This raises questions about the unit of analysis for evaluations cross-programme? cross-OP? and how to plan integrated evaluation programmes

Rigorous Impact Evaluation is Difficult

- Top-end Impact Evaluations moves evaluation to the frontiers of sophisticated, expensive and still-developing research methods in the social and economic sciences

- Methods such as propensity score matching, discontinuity designs and instrumental variables (on the counterfactual side); and QCA, Contribution Analysis, Process Tracing and Realist Evaluation in relation to Theory Based Evaluations are not simple

- There are few specialists in most EU Member States able to apply them

- This is why we need to be cautious and selective about when we would expect to do this work

There are more modest ways of evaluating results

- Sometimes we can rely on 'descriptive inference' – an accumulation of different kinds of evidence about change occurring in some programmes but not in others without the same interventions

- If for accountability purposes we want to be able to say that a) there has been an improvement in results and b) that it seems to be caused by a programme then knowing *how* the programme worked (explanatory analysis) may not be needed

- Sometimes we can rely on prior evaluation, research and theory – we do not need to 'prove' again that smoking is bad for your health or that nutritional food is good for children! Here we might simply focus on implementation and uptake

Engaging stakeholders, beneficiaries and others implicated in evaluations

Validity is a major 'quality' criteria in any evaluation

- Systematically involving users, stakeholders, beneficiaries, panels of citizens, policy makers is a good way to improve validity

- This needs to take place at every stage in the evaluation process – from operationalising criteria – e.g. deciding what outcome descriptor or measure represents 'success'; to interpreting data when drawing conclusions and making recommendations

These things are too important to leave to evaluators!

Conclusions

- Evaluations need to be designed to take account of Evaluation Questions; the characteristics of programmes; and the capabilities of available methods

- Evaluations often use different designs to answer different EQs; & use different methods because all have their strengths and weaknesses

- IEs variously need to establish causality; explain why and how; and to assess how programmes interact with other causative factors- however IEs and results are not just about causl & explanatory designs

- Deciding on the 'unit of analysis' for an evaluation – sites, programmes, multiple programmes - is also part of the design process

- Validity will be increased if users, stakeholders, beneficiaries and citizens are involved in evaluative design and evaluative judgements