



Delivering Evaluations in Uncertain Times

Elliot Stern, Emeritus Professor of Evaluation Research, Lancaster University

10th Evaluation Conference, 12-13 November 2024 Prague

Disclaimer

- Invited here as an expert working with the Evaluation Helpdesk, a body that supports evaluation for both DG REGIO and DG EMPL
- I will be drawing on facts and experiences that come from working with the Helpdesk over many years
- However, opinions and judgements about '*what is (not) going well in evaluations and why*' are mine alone – personal reflections that should not be attributed to the Helpdesk

Helpdesk Background

- Evaluation Helpdesk in present form established in 2015
- Primary purpose of the Helpdesk to improve evaluation quality of Cohesion Policy evaluation across the EU – ultimately to improve CP programmes
- The Helpdesk is led by 3 consulting companies: *Applica srl* (Brussels) *Ismeri Europa* (Rome) & *Quadrant Conseil* (Paris) supported by an internal team and a network of consultants and experts
- Similar but less comprehensive support for evaluation in REGIO since 2011 - Ad hoc support and peer review activities in DG REGIO since early 2000s

Main Helpdesk Activities -1

- Review and assess all CP evaluations published by Member State
 - Maintain an online library of these reviews - 3,000 evaluations published by MS reviewed and summaries input into Evaluation Library
- Review and assess MS evaluation plans in terms of objectives, design and data, organisation, expertise of evaluation units & plans for evaluation use
 - Support DG REGIO with evaluations they commission

Main Helpdesk Activities - 2

- ‘On-demand’ support for MSs & Managing Authorities - on ToRs, report quality, improving methods, evaluation ‘supply’ approaching 70 missions in last 5 years
 - Most are face to face but some online
- Ad hoc training/development events – e.g. for newly recruited Managing Authority staff; and beginning to pilot ‘regional network meetings’ so MSs can more easily share experience
- Organise a Summer School for MSs lasting 2 to 3 days
- Help organise biennial REGIO Evaluation Conference

So, what 'goes well' these days?

General but uneven evolution.....

- From conformance checking and concern to 'spend the budget' to focusing on effectiveness, results & strategic goals
- From 'keeping Brussels happy' to seeing evaluation contributing to development of national and territorial policy
 - From MAs designating generic civil servants as evaluation 'specialists' to small cadres of capable evaluators and evaluation managers in many MAs

Many MAs now take a longer-term perspective

- Invest in capacity development
- Have evaluation strategies/policies
- Greater interest in evaluation 'use'
 - More staff in post although...

still meet 'ghost evaluators' who are designated as such but have other jobs in planning, statistics and general administration

A 'mixed-bag' in terms of evaluation quality....

- If we operationalise 'quality' as *quality control* – how good are the final reports- a higher proportion nowadays would be judged as 'good enough' in technical terms
 - In these reports initial questions are answered; acceptable methods used; and evidence supports conclusions
- This would certainly be true for evaluations of less complex OPs or policy interventions – active labour market & business support where counterfactuals may be appropriate

Methodological weaknesses are widespread

- However counterfactual evaluations often fail to identify suitable control groups – where control group *behaviour* sufficiently resembles the supported group about businesses *overall* or employment effects *in general* (external validity)
- Reliance on any single method will inevitably makes it difficult to draw general conclusions....
- Theory-Based evaluations that should help explain and generalise are also often flawed – in most TBEs reviewed by the Helpdesk causal pathways are inadequately mapped & causal mechanisms together with enabling/risk factors not well-developed

Systemic weaknesses

- If we operationalise ‘quality’, as *quality assurance* i.e. fitness for purpose in relation to Cohesion Policy, rather than *quality control* of reports, the picture is also inconsistent
- But here we are looking at systemic problems – not those that can be reasonably attributed to poor management or planning in any one MA or evaluation team
 - I want to briefly outline 6 such ‘systemic’ weaknesses

1. Timing of evaluations and of policies do not align

- Some systemic problems are well-understood, as when evaluations are poorly synchronised with the policy cycle. E.g.,
 - Programme results span 2 or 3 programming periods, something EU evaluations are not set up to address
 - In the shorter term when ‘impact’ evaluations come in after next programming periods have been decided
- These problems are remediable – commission longer-term evaluations!
- However, at present many evaluations are untimely or not used because of poor synchronisation

2. Not distinguishing between policy failure and implementation failure..

- MA evaluations are very focussed on results and impacts despite the time-lags that weaken their explanatory claims
- They are very weak evaluating implementation – except in terms of procedural compliance, effective contracting, and timely expenditure
- This is partly because the implementation process for a growing proportion of ERDF interventions are outside the MA's 'span of control' relying on partners, delivery agents, subcontracted projects and territories
- For whatever reason this constitutes a major systemic weakness

3. Lack of attention to related programmes both European and national

- ‘Owners’ of policies are always keen to evaluate their own programmes and policies
- In the real-world policies intersect – national, regional and multiple EU initiatives target the same ‘beneficiaries’ or their ecosystems
- All MSs benefit from a variety of EU programmes such as Horizon and RRF and as MSs become less dependent on EU funding and more on own-resources national programmes are also difficult to separate out.....
- But adjacent policy programmes are only rarely evaluated

4. Poor Institutionalisation in Government

- Evaluation of Cohesion Policy and Structural funds are part of a country's machinery of government – effective administrations are more likely to have effective evaluation systems
- There are outliers – there are effective evaluation functions in unstable bureaucracies and weak evaluation functions in well-organised bureaucracies; but these are exceptions
 - Effects can be severe when for example
- Evaluation functions in MAs are not given enough staff; MAs lack support from major ministries; and procurement is problematic when awarded evaluation contracts are challenged in court

5. Focussing on Government not Governance

- When most ministries and agencies think about stakeholders and potential users of evaluation, they mainly consider those in government – Ministers and civil servants
 - RTD might be an exception here, when the role of Large firms, SMEs, Universities and Research Institutions cannot be ignored
- However contemporary ideas about governance with its extended notion of decision making/influence are not routinely considered in CP/ERDF evaluations
 - This limits evaluation's ability to communicate with all relevant actors and reduces the potential learning benefits of evaluation, often in spaces where learning could be most useful

6. The challenge of complexity

- Evaluation technics and practice are not well-adapted to complex policies – i.e. policies where multiple causal mechanisms are mobilised to achieve ‘composite’ outcomes in diverse, non-standard contexts.
- Contemporary policies are increasingly complex partly because of complex policy ambitions – to reduce inequality, promote innovation, balance conflicting priorities, ensure green transitions and sustainability....
 - Policies are also complex because the implementation and delivery arrangements needed for contemporary policies involve many agents, partners, delivery agents, new forms and extended time horizons

From policy delivery to policy uncertainty

- As anyone who has worked with Smart Specialisation and Mission Driven Policies will know, evaluation is itself transformed when the ‘objects’ of evaluation are uncertain, distributed, constantly changing and coordinated by evolving governance institutions
- What do we mean by ‘high quality evaluation’ when evaluation becomes reflexive, collaborative, decentralised, formative and multidisciplinary?

Emergent Policy Scenarios

- Finally, I'd like to step back from specific 'systemic' weaknesses that influence current limitations of CP and ERDF evaluations more generally
- Thinking in terms of existing and emergent policy scenarios allows us to anticipate new threats to evaluation quality and relevance unless the implications of these scenarios for evaluation are anticipated

New policy scenarios shaping evaluation

- We are a long way from the stable policy scenarios of the ‘service delivery state’ or even the ‘enabling state’
- The acknowledged mainstream policy scenario in EU as well as in Member states, is *Transition Management* driven by sustainability, energy and biodiversity concerns
 - Another scenario less publicly acknowledged is *Crisis Management*. Over the last 15 years we have seen consistent disruption by crises – the financial crisis, Covid-19, geopolitical disruptions and of course climate change

Policy uncertainty meets risk aversion

- Both these scenarios are characterised by high levels of uncertainty and risk, however our policy systems – and the culture of policy-making – is often intolerant of uncertainty & risk averse
 - *Transition Management* relies on local and decentralised experiments and behaviour change as much as top-down fiat; while *Crisis Management* relies on decentralised collective action and consensus building/conflict resolution

Evaluation within new policy scenarios

- The evaluation community has some – but not all- of the toolkits, models, ethical frameworks and practice norms to contribute to these new scenarios – I’m thinking here of recent experience with policy labs, citizen science and participatory ‘self-evaluation’ models
 - However as the evaluation community has itself become ‘institutionalised’ it is also more accustomed to working with top-down policy assumptions and less skilled working with local actors & experiments or bridging different tiers of multi-level governance

Policy systems are also adapting....slowly

- Centralised policy-systems are also slowly coming to terms with adaptive policy making and planning; with horizontal/interministerial working in government; and with foresight as well as hindsight
 - But most central policy systems still struggle to accept uncertainty as the norm – demanding ever more ‘rigorous’ answers to evaluation questions; resisting innovation from the periphery; and finding it difficult to engage citizens as partners in innovation let alone in evaluation

In Conclusion- and to end on an almost optimistic note!

*New policy scenarios of ‘transition management’ and ‘crisis management’ will bring new systemic threats to evaluation quality
.....more things that might not always ‘go well’*

*If we think through the consequences of these policy scenarios,
many threats to evaluation quality and relevance can be made
manageable – risks to evaluation quality can be mitigated if not
entirely eliminated*

Thank you for your attention!