










Thoughts on the future of policy evaluation from the perspective of an evaluation contractor

Paul Simmonds

**CECAN Looking Back and Looking Forward:
Closing the loop from evaluation to policy making**



Recent trends we expect to continue shape policy evaluation

-  Further **embedding** evaluation within the policy cycle
-  Expand evaluation for **learning**
-  Developing evaluation methods for societal impacts
-  Improving counterfactual analysis
-  Improving our ability to summarise evaluation findings
-  Improving data infrastructure
-  Automation / AI



Further embedding evaluation within the policy cycle

- ↗ Evaluation is less incidental and better for it
 - ↗ There is still room for improvement: organisation-wide evaluation strategies
- ↗ We expect policy makers to continue to expand their evaluation work for reasons of public accountability and for learning
 - ↗ We have seen the **re-emergence of process evaluations**
 - ↗ An interest in **policy experimentation**
 - ↗ And more systematic reviews / **meta-analyses** of what works (looking at evaluations at scale)
- ↗ We expect Government to maintain its demand for evaluation
 - ↗ Mostly purchased from independent professional evaluators,
 - ↗ But with more routine evaluations will be delivered in-house



Developing evaluation methods for societal impacts

- ↗ While the need to capture economic impacts is unlikely to diminish
- ↗ It may become more critical in some aspects (e.g., SR / BC templates focus on hard figures only)
- ↗ We expect relatively greater emphasis to be placed on efforts to better capture and value non-economic impacts of public interventions
 - ↗ Community
 - ↗ Natural capital
 - ↗ Culture



Counterfactual analysis in complex settings

- ↗ The increasing focus on societal challenges and complex interventions in complex, open systems has challenged conventional approaches
- ↗ Still a work-in-progress
 - ↗ Combining TBE based counterfactual methods with QED
 - ↗ Running multiple counterfactuals
- ↗ In simple terms, we are repeatedly asking the questions: has it changed; and has it changed more than it might have done anyway
- ↗ Combining many partial views



Summarising and monetising ...

- ↗ With interventions in complex systems, policymakers are looking to deliver improvements on many fronts
 - ↗ Community engagement
 - ↗ International visibility
- ↗ One or two are readily monetisable, most are not
- ↗ As evaluators, we still need to weigh many types intangibles to understand effectiveness and value for money
 - ↗ Contingent valuation methods are one line of development that can result in a single, financial indicator (RoI), but not always credible
 - ↗ Multi-criteria analysis is an alternative approach which summarises overall performance and provides a basis for comparing options or benchmarking with similar interventions
 - ↗ Increasingly used together: a partial RoI and a fuller MCA



Improving data access and data infrastructure

- ↗ The gradual improved access to contextual data is one of the most significant developments in recent times
 - ↗ Monitoring systems have improved
 - ↗ Commercial databases have grown in number and coverage
 - ↗ Government and other open-access datasets have also expanded in number
- ↗ We see a shift in the balance of primary and secondary data collection, and a concern to minimise burden
- ↗ We see this as an area for major future development



Automation / AI

- ↗ Evaluators have seen substantial process automation and digitalisation over the past 25 years
- ↗ AI may deliver a similar scale of change in the next 5-10 years
 - ↗ 100s of tools are being developed that will make individual tasks quicker and easier (e.g., a preliminary summary of a technical report)
 - ↗ Data linking and matching (e.g., using fuzzy matching beyond unique IDs)
 - ↗ For evaluation more directly, AI should transform our ability to quickly interrogate and synthesise large volumes of partially structured and qualitative data (e.g., classifiers of textual data) and re-analyse those data reasonably quickly where priorities change
- ↗ Quite where it will take us is unclear, but
 - ↗ Evaluators are investing heavily
 - ↗ Large numbers of tech platforms are doing likewise
- ↗ Ultimately, evaluators should be able to do more and more quickly
- ↗ However, we should expect a period of experimentation and learning among both clients and contractors, as the techniques develop and gain provenance



An important systemic change that we expect to see continue and maybe increase

- ↗ Evaluation as an accompanying measure, running alongside the intervention across its lifecycle
 - ↗ Multiple evaluation work packages
 - ↗ Multiple reporting points
 - ↗ A more participatory approach



Other classic ‘contractor’ challenges

- ↗ Measuring impact early
 - ↗ Funding cycles make it necessary to consider ultimate impacts before they occur
 - ↗ TBE methods are an important partial response
 - ↗ More space needs to be given to estimating likely future effects
 - ↗ Plus, interventions should be looked at more robustly some years later when impacts can be determined more completely
- ↗ Staff turnover within client teams
 - ↗ It is not unusual to see client teams change several times within the course of a single evaluation
 - ↗ This can be disruptive particularly where a new team wishes to take things in a slightly different direction
 - ↗ It can also mean the new team has a steep learning curve
 - ↗ Evaluators could be briefed on changes before and ideally feed into the handover



Other contextual issues to reflect upon

- ↗ Policy interventions with unrealistic ambitions (**smallness**)
 - ↗ Tight public finances can lead to interventions that are too small
 - ↗ Difficulties in identifying material impacts seen as failures of measurement
- ↗ Relatedly, we are often asked to evaluate an intervention that is one of many other small initiatives with a broadly similar brief
 - ↗ Crowded landscapes demand greater effort to test coherence and additionality (strong policy messaging can lead to **fragmentation**)
 - ↗ It can mean working closely with wider actors as subjects of interviews but also as contributors to research and sample design



And should we begin to simplify complexity?

- ↗ A TBE approach is well able to deal well with complexity
 - ↗ On the downside, a comprehensive TBE can be time consuming and costly, and some level of simplification may make it more affordable
- ↗ We see ToCs becoming more elaborate; 10-20 has become 30-50 building blocks, with detailed risks and assumptions
- ↗ These more complete models better reveal all the connections (reducing the gap between outcomes and impacts)
- ↗ But the level of detail can overwhelm evaluations (completeness, cost)
 - ↗ Breadth: lightly test the progression across all chains of cause and effect
 - ↗ Depth: deep dive programme contributions to key outcomes



Real-world evaluation in a complex world

- ↗ I have talked a lot about individual policy evaluations however most clients have to think in terms of many evaluations not one; and these must be carried out within a SR period, fit an overall budget and reflect the in-house capacity available to specify and manage
- ↗ A portfolio approach may be way forward
 - ↗ Focus on the larger / critical interventions and the novel
 - ↗ Sample the smaller and more routine interventions
- ↗ Proportionality: scale, complexity, novelty ...



Thank you.

Abidjan • Amsterdam • Berlin • Bogotá • Brighton • Brussels • Frankfurt/Main • London •
Paris • Stockholm • Tallinn • Vienna