Using an assessment of 'complicated' and 'complex' characteristics to determine evaluation design of innovation policies

Abstract for Open Evaluation 2016 conference

Title: Using an assessment of 'complicated' and 'complex' characteristics to determine evaluation design of innovation policies

Author: Jonathan Cook, Director at SQW Ltd

Address: SQW, 1st Floor 43 Chalton Street, London NW1 1JD, United Kingdom

Phone number: +44 (0) 20 7391 4105 or +44 (0) 7715 071501

Email: jcook@sqw.co.uk

Track: Relates to the conference theme of "RTI policy evaluation in the policy-making process", with a focus on the "specific characteristics of RTI as determinants of evaluation approaches".

Keywords: evaluation design, impact evaluation, programme theory, theory-based approaches, complexity, RTI infrastructure, R&D grant policies, demand-side innovation policies

Abstract

Context and objectives

Policies in the field of Research, Technology and Innovation (RTI) are often not associated with linear processes of cause and effect. This partly reflects the nature of innovative activities, whereby the results and the ways in which firms apply these can be unclear at the outset, in particular where innovation takes place at the frontiers of knowledge (Hof *et al.*, 2012). In addition, the increasingly 'open' and collaborative way in which innovation is undertaken can mean that some of the benefits of RTI policies are indirect and unintended, as results are diffused through the innovation network system (Jordan, 2010), e.g. through knowledge spillovers. Moreover, the benefits of RTI policies can be unevenly distributed, with small numbers of beneficiaries/actors reaping the vast majority of the rewards (Cook *et al.*, 2013).

The specific ways in which some RTI policies are designed create challenges for evaluators seeking to assess cause and effect. For example, reflecting the iterative and collaborative process of innovation, policies can involve multiple components or partners. Other policies may involve support that is highly tailored to specific contexts and circumstances such that no 'standard' intervention exists. The rise of 'demand-side' policies has also resulted in the need to consider a broad set of inter-relationships between different institutions within a system. Alongside these challenges, policy-makers' expectations of evaluation are high. In particular, there is a desire for evaluation to place a 'value' on policies and programmes to

inform future decisions and investments, ideally through the use of experimental or quasi-experimental approaches. How can evaluators best respond to these challenges and expectations?

Evaluation literature has suggested that the characteristics of interventions can be used to inform evaluation design, with Rogers (2008) drawing a distinction between aspects of interventions that can be categorised as 'simple', 'complicated' and 'complex'. Rogers (2008) also illustrated how programme theory can be used in complicated or complex situations without resort to "messy" logic models. For interventions that exhibit features that are complicated or complex (or where the population of beneficiaries is small), counterfactual-based approaches to evaluation may be inappropriate. In these cases, theory-based evaluation approaches may be used (Rogers, 2007; Weiss, 2000; White, 2009). White and Phillips (2012) described a range of these theory-based techniques, including contribution analysis, process tracing and realist synthesis, which can be used to assess the extent to which interventions have brought about outcomes.

In this paper, we draw on our recent evaluation studies relating to RTI policies to examine the extent to which different policy interventions exhibit the 'simple', 'complicated' or 'complex' characteristics set out by Rogers (2008). We then describe how these characteristics can be used as determinants of appropriate evaluation design, and the role of programme theory as a tool to inform evaluation. Finally, we examine the extent to which evaluation designs for RTI policies with 'complicated' and 'complex' characteristics are likely to meet policy-makers' expectations of valuing the contribution of RTI policies to the economy.

The paper draws on the experiences of several recent evaluation studies that we have undertaken covering evaluation scoping studies and programme evaluation assignments. They are focussed on business innovation in the UK and the EU, and the mix of interventions includes: single company R&D grants; collaborative R&D grants; investments in new RTI infrastructures that seek to bridge the gap between research and businesses; and demand-side innovation policies.

Assessing the characteristics and programme theories of policies

Drawing on the classification of issues identified by Rogers (2008), Table 1 sets these into the context of RTI policies, and extends the range of issues to cover other aspects that we have found to be important. In summary, these aspects are as follows (drawing on Rogers, 2008):

- The nature of implementation and engagement takes account of the extent to which multiple partners are involved in delivery or as part of innovation partnerships.
- Simultaneous causal strands mean that two or more routes to outcomes are required to occur for an intervention to work, such as technical success of an R&D project along with the development of innovation capacities to take the output to market.

- Alternative causal strands are subtly different as they mean that there could be more than one causal route for a programme/policy, which can be particularly relevant where the intervention is highly tailored.
- Timescales to outcomes can be long for RTI policies, and a further issue is the extent to which they may vary across a policy, e.g. with recipients of an intervention achieving outcomes over different timeframes.
- Policy objectives may be focussed on economic issues (e.g. related in some way to growth), though may also cut across a range of issues (e.g. economic, societal and system).
- Non-linearity of outcomes reflects that an initial effect may result in a feedback loop that brings about further rounds of effects, i.e. acts as a tipping point.
- Outcomes can be pre-identified based on known or at least anticipated relationships (i.e. non-emergent outcomes) or can be dependent on interactions between different organisations, and sometimes be unpredictable (i.e. emergent outcomes).

Table 1: Complicated and complex aspects of RTI policies

Aspect	Simple version	Non-simple version	Examples from RTI policies of non-simple version
Nature of implementation and engagement with the policy	Businesses/ organisations benefit on an individual basis from the policy	Multiple partners are involved when businesses/ organisations engage with the RTI policy (Complicated)	Collaborative R&D schemes RTI infrastructure Demand-side measures
Simultaneous causal strands	Single (at least primary) causal strand	Multiple causal strands (Complicated)	Various RTI policies, for example: specific outcomes relating to an R&D project supported by a policy (e.g. progress through technology readiness), alongside other outcomes such as development of innovation capabilities and feeding back into the research base
Alternative causal strands	Experience of the policy is the same/ similar, with broadly the same causal mechanism	Different causal mechanisms depending on context	RTI infrastructure: businesses' experience can vary, for instance as they select the support that meets their needs;

Aspect	Simple version	Non-simple version	Examples from RTI policies of non-simple version
		(Complicated – where variation can be categorised/ coded; Complex – where experience essentially bespoke)	feedback loops may also result in refining existing or bringing about new R&D projects
Timescales to outcomes	Same/similar for those benefiting from the policy	Variation in timescales to outcomes, e.g. reflecting technologies and markets (Complicated)	Collaborative R&D schemes and RTI infrastructure: the timescales to commercial benefits for businesses potentially vary from under 5 years to 15/20+ years
Policy objectives	Single, e.g. focussed on economic	Multiple, e.g. combination of economic, societal and system (Complicated)	Demand-side measures, where the purpose is to bring about economic growth, domain-specific objectives (e.g. clean energy) and change within the innovation system
Non-linearity and disproportionate outcomes	Linear causality and proportional impact	Feedback loops and the potential for a critical tipping point to bring about a large ultimate effect (Complex)	Demand-side measures, where small initial effects (e.g. increased initial take-up) can lead to a large ultimate effect (e.g. through feedback to innovators and wider diffusion)
Emergent outcomes	Outcomes can be pre-identified, e.g. increased R&D spend and business performance metrics of those directly involved	Outcomes dependent on the interactions between organisations, and how the behaviours are influenced (Complex)	Demand-side measures where effects rely on system changes, such as the interaction between different organisations to create appropriate frameworks

Source: Author, drawing on Rogers (2008)

This classification provides a set of determinants for evaluation approaches. For example, the proliferation of characteristics that are 'simple' will lend weight to experimental or quasi-experimental approaches, and where 'complex'

characteristics are significant, theory-based (or alternative) approaches will be required. For policies with 'complicated' characteristics, there may be a choice or a mix of experimental/quasi-experimental and theory-based approaches, depending on other parameters and key evaluation questions.

The final choices of evaluation design will then require further consideration and refinement. For example, an R&D grants policy may have mainly simple characteristics (e.g. on implementation and engagement with the policy, causal strands, and policy objectives), and a quasi-experimental approach can be adopted. However, the high degree of skewness in outcomes, the potential variability in timescales to outcomes, and the alternative policies available in the wider RTI landscape can pose challenges to analysis, requiring triangulation between methods (e.g. see SQW *et al.*, 2015). Moreover, spillovers are relevant to this RTI policy, and these require some form of case-based research that seeks to track through how the original intervention has contributed to these effects. In essence, therefore, for most RTI policies, a single evaluation approach is unlikely to yield satisfactory findings.

For evaluators, developing a sound classification of these issues can require in-depth research with those involved in delivering the RTI policies. This participative approach should help evaluators to develop programme theories that better reflect the realities of policies, and ultimately evaluation design that is more appropriate and sensitive to the characteristics of policies and how they are implemented.

There is a risk that the logic models and programme theories for interventions with complicated and/or complex characteristics become too "messy", with every box in the logic seemingly linked in some way to every other box (Rogers, 2008). An alternative is to adopt a common structure or framework, within which a series of bespoke theories and logic models (or sub-theories) can be developed for individual projects or 'cases' (Cook, 2016; SQW and Cambridge Econometrics, unpublished). These sub-theories are particularly appropriate where there are numerous alternative causal routes, e.g. because an intervention is highly tailored in its implementation to particular contexts. This approach also facilitates building evidence on recursive feedback loops and interactions that are emergent, because it is designed to be easily refined and additive. Figure 1 provides a stylised example, illustrating three steps that show how more complicated aspects can be incorporated and how the approach can be tailored to different contexts:

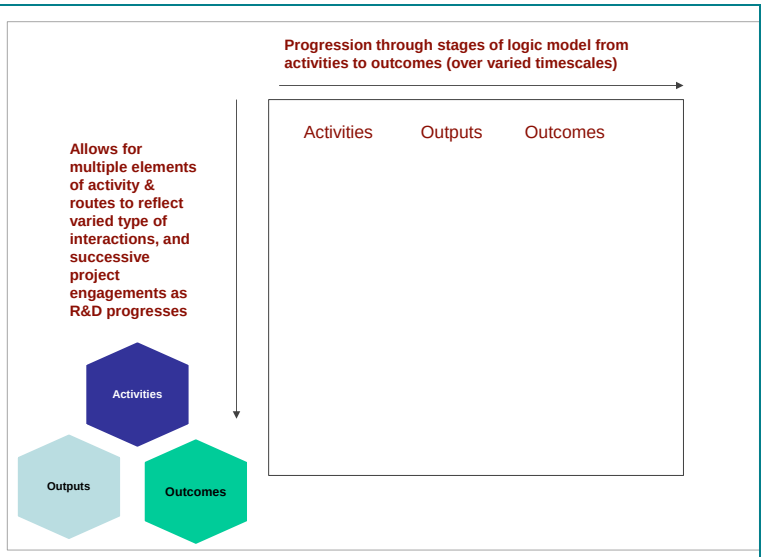
- the overall structure is shown in step 1
- the population for a relatively simple intervention is shown in step 2
- and how this might be used to consider alternative or simultaneous causal strands and also new interactions/activities is shown in step 3.

Figure 1: Using flexible programme theories to facilitate evaluation

Step 1

All activities, outputs, and outcomes are allocated an individual 'tile' to be included on the logic model 'slate'.

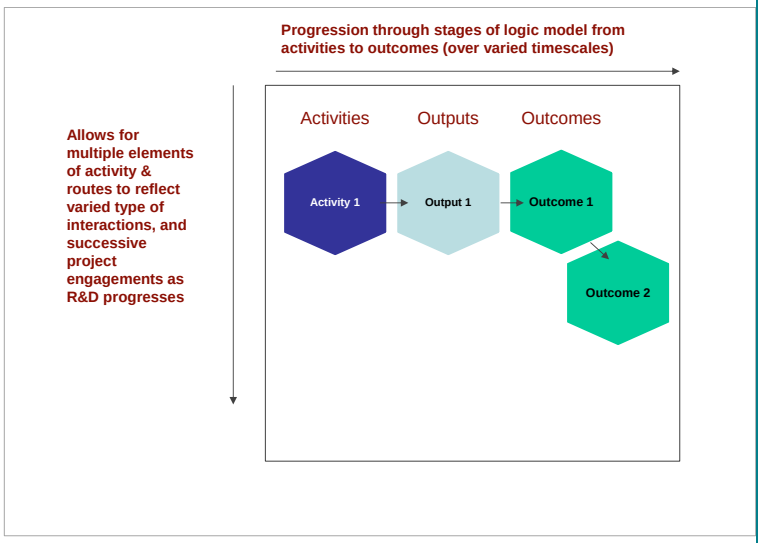
The 'slate' has two axes: a horizontal axis setting out the progression from activities through to outcomes (for which an underpinning theory of change is developed), and a vertical axis, that enables multiple forms of support or outputs/outcomes to be captured within the logic model.



Step 2

The logic model 'slate' is populated with the relevant activity/activities, with the expected outputs and outcomes flowing from these identified and included – and a postulated theory of change developed.

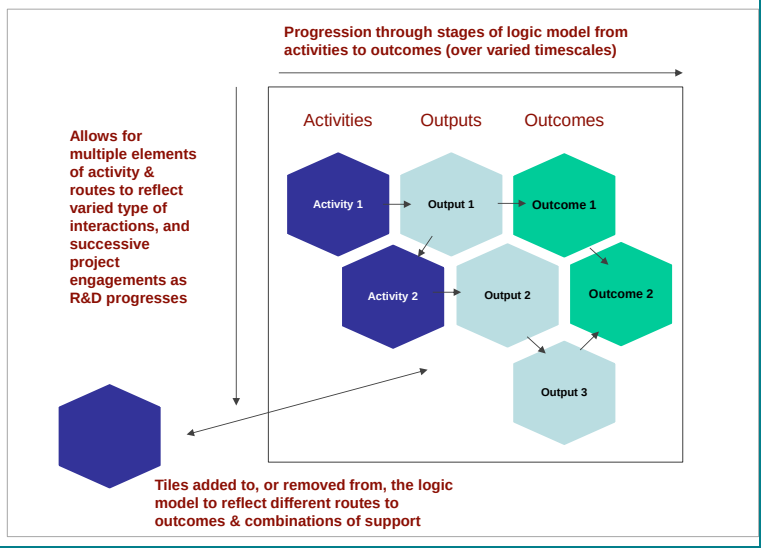
The approach opposite sets out a relatively simple generic logic model where one activity is provided, leading on to a single output and two outcomes



Step 3

The logic model 'slate' is updated through the addition of more 'tiles' based on evaluation evidence (e.g. if further activities, output, outcome combinations are identified) or to reflect increasingly complex and complicated routes to outcomes.

The example opposite sets out a theory of change where two activities delivered in parallel lead to a range of outputs and outcomes.



Source: Author, drawing on SQW and Cambridge Econometrics (unpublished)

Using a theory-based approach

As indicated in Table 1, most of the RTI policies examined here have complicated and/or complex characteristics, which pose challenges to evaluation. Responding to this, our work to scope evaluation approaches has frequently drawn on these determinants in recommending mixed methods within an overarching theory-based approach (e.g. SQW and CE, 2016). For example, contribution analysis is one such approach that can be taken, i.e. examining whether there is strong evidence that the intervention, rather than other factors, was critical in causing the outcomes observed (distinct from evaluating what would have happened in absence of the intervention). In the case of RTI policies such as collaborative R&D schemes and investments in RTI infrastructure, this is likely to involve collating a range of evidence in order to test, from different perspectives, the contribution of the policy under examination. This may include evidence from:

- project-specific case studies and beneficiary interviews to test, bottom-up, the contribution of the intervention to outcomes
- interviews/case studies with indirect beneficiaries to test the extent to which spillovers may have been achieved, and how far these relate back to the original intervention
- technology mapping combined with interviews with sector experts to assess, from a top-down perspective, the contribution that an intervention has made to more systemic change or technology development.

Such approaches can provide, in a transparent way, an assessment of whether, how and in what context, RTI policies have brought about their intended outcomes and also unintended outcomes. However, the extent to which the outcomes can be quantified and monetised will be limited, even at the level of individual beneficiaries of policies, let alone at the level of the policy overall. This may leave unanswered the policy-maker's question relating to the value of the policy. In some cases, a partial assessment may be possible here. Again, relating back to the classification of the characteristics in Table 1**Error! Reference source not found.**, for RTI policies or for the aspects of RTI policies that are merely (!) complicated, and for which a single or small number of key outcomes can be observed or assessed, a quasi-experimental approach could be used. Therefore, for at least part of the policy, some value can be ascertained. There is an important communications issue here, which relates to the tendency for audiences of evaluations to focus on what can be counted. Given the potential for RTI policies to lead to spillover effects and disproportionate outcomes that cannot be quantified, there are risks that these receive insufficient attention, thereby under-stating the effects of the policies, and creating perverse incentives for implementation to focus too much resource on the more direct routes to outcomes.

Learning points

Several key lessons are relevant for evaluators, policy-makers and deliverers involved in RTI policies. First, the characteristics of interventions can be important determinants of evaluation approaches. However, classifying policies

by these characteristics is not always neat and straightforward, because policies may often have combinations of simple, complicated and complex features. Second, there is a need to develop appropriate programme theories and logic models, which particularly draw on the perspectives of those delivering on the ground. Third, for interventions with complicated or complex characteristics theory-based evaluation approaches provide an important option where counterfactual-based approaches are inappropriate or to complement counterfactual-based approaches. Finally, even where parts of RTI policies can be evaluated using counterfactual-based approaches, policy-makers need to be alert to the partial story provided and the potential for such results to distort behaviours.

References

Cook, J. (2016) 'Making the infinite countable? Responding to the challenges when evaluating innovation policy', *Presentation to the UK Evaluation Society Conference 2016*

Cook, J., Macdonald, B. and Pates, R. (2013) 'The Pareto Principle: the importance of the vital few in business support programmes', *SQW Insight Paper*

Hof, B., Koopmans, C., Lieshout, R. and Wokke, F. (2012) 'Design of a Methodology to Evaluate the Direct and Indirect Economic and Social Benefits of Public Investments in Space', *SEO Economic Research Technical Note*

Jordan, G. (2010) 'A Theory-Based Logic Model for Innovation Policy and Evaluation', *Research and Evaluation*, Vol.19 No.4

Rogers, P. (2007) 'Theory-based evaluation: Reflections ten years on', *New Directions for Evaluation* 114: 63-81

Rogers, P. (2008) 'Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions', *Evaluation*, Vol. 14 No. 1

SQW, Cambridge Econometrics and BMG Research (2015) 'Evaluation of Smart', *Report to Innovate UK*

SQW and Cambridge Econometrics (2016) 'The Aerospace Technology Institute: Scoping study to establish baselines, monitoring systems and evaluation methodologies', *BIS Research Paper* No. 271

SQW and Cambridge Econometrics (unpublished) 'High Value Manufacturing Catapult Impact Evaluation Framework', *A Report to the HVM Catapult*

Weiss, C.H. (2000) 'Which links in which theories shall we evaluate?' *New Directions for Evaluation* 87: 34-45

White, H. (2009), 'Theory-based impact evaluation: Principles and practice', *Journal of Development Effectiveness* 1(3): 271-84

White, H. and Phillips, D. (2012) 'Addressing attribution of cause and effect in small n impact evaluations', *3ie working paper*, June 2012