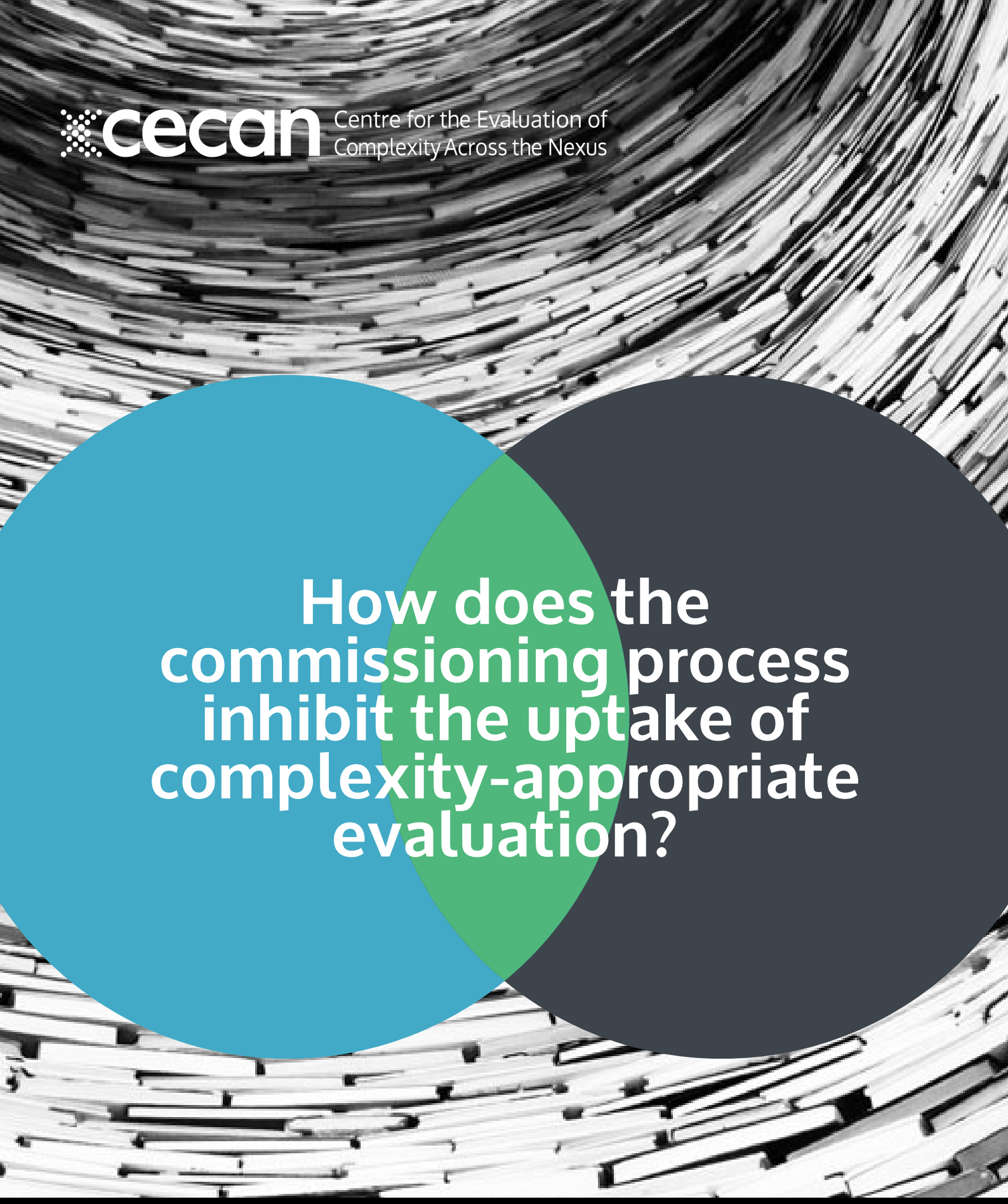

How does the commissioning process inhibit the uptake of complexity-appropriate evaluation?

Jayne Cox
Director, Brook Lyndhurst

Supported by CECAN mentors
Pete Barbrook-Johnson, Ian Christie, and Emma Uprichardt

Summary

CECAN has been working with policy teams and social scientists in UK government to explore how 'complexity-appropriate' evaluation can enhance the usefulness of policy evaluation.

Complexity-appropriate evaluation can be characterised as an approach based on ways of thinking drawn from complexity science and systems theory. It emphasises adapting to emerging findings, involves iteration and multi-stakeholder working, and uses methods that can capture the full complexity of the policy and context being evaluated (e.g. path dependency, emergence, feedback loops, multi-causality etc.).

These approaches are in direct contrast to conventionally linear ways of framing and conducting evaluation studies. If they are to be used more widely in policy evaluation they will need to gain traction in externally commissioned evaluations. In recognition of this, CECAN supported a study to investigate the role of evaluation commissioning in helping or hindering the take-up of complexity-appropriate evaluation methods.

This paper presents findings from interviews with 19 commissioners and contractors who have been involved in specifying or responding to tenders for policy evaluation studies, mainly for the UK Department for the Environment, Farming and Rural Affairs (Defra).*

The study focused on how approaches and methods are selected in the competitive tendering process for policy evaluations; and how the wider operating context of the individuals involved in the process can influence their behaviours during the commissioning process.

The context for policy evaluation is constantly evolving. Sometimes opposing pressures to raise the quality of evaluation practice and to "do more with less" (budget) tend to favour risk averse behaviours and sticking with well-known and accepted methods (e.g. counterfactual impact methods). Equally, some respondents are finding that "standard" methods are not suited to the complexity of some of the questions they are being asked and would welcome a more plural "evaluation toolbox".

The rules and processes of competitive tendering can, however, inhibit methods innovation. There is rarely enough time or flexibility in the tender specification process to consider less well-known alternatives; and the tender assessment and scoring process was highlighted as problematic. Respondents said it is unsuited to selecting effectively between radically different approaches and tends to favour quantity over quality of inputs, and tangible over intangible value in contract 'deliverables'.

In addition to barriers to methods innovation generally, respondents thought the tendering process deals poorly with two of the intrinsic features of complexity-appropriate approaches – the need for flexibility and iteration to respond to emergence, and the value of collaborative working and co-creation.

Many suggestions were made for changes within existing commissioning processes that could help contractors and commissioners to specify complexity-appropriate methods, including:

- more scoping studies and meaningful contractor dialogue upstream from tender specifications;
- finding ways to accommodate uncertainty and flexible evaluation designs in tender specifications and scoring (plus a price guide as standard) and to not penalise bids that include them;
- measures to enable more responsive, and risk-aware, project and contract management.

Building demand from policy end-users would also be essential, including challenging linear ways of thinking about policy and outcomes. Three components were highlighted: more effective knowledge exchange; upskilling and fostering a community of practice of commissioners and contractors; and championing of complexity-appropriate evaluation by individuals with influence, power and access to higher levels of policy. Respondents also identified a wide range of questions that would need to be answered in a 'business case' for promoting complexity-appropriate methods. Evidence from the CECAN case studies could be used to provide answers.

The most critical interviewees, however, suggested a more radical overhaul of research procurement processes is needed at a higher level. They want procurement to be simpler, quicker and more flexible. Change would require a lead from procurement functions in government, to explore alternative commissioning models that would enable more flexible and collaborative approaches to tendering and contract management.

* **Important note:** the findings reflect the views of the individuals taking part and do not represent the views of Defra or the contractor organisations.

1 | Background and aims

With CECAN's support, policy teams and social scientists in UK government have been looking at new ways to think about and evaluate policies through the lens of complexity[1], which are in direct contrast to conventionally linear ways of framing and conducting evaluation studies. They have applied a wide range of 'complexity-appropriate methods' to policy evaluation case studies[i] (e.g. realist and theory-based approaches, systems mapping, qualitative comparative analysis, process tracing, agent based modelling), learning how to apply the methods and how that could enhance the value and outcomes of policy evaluation.

The next step for these methods is for them to become commonplace in government evaluation practice, which means in externally commissioned evaluations as well as internal projects. Competitive tendering is one of the main channels through which new approaches and methods can break into policy evaluation and be mainstreamed, so it is worthwhile to examine how those procurement processes work to identify where the barriers and opportunities are for adopting complexity-appropriate methods.

Evaluation commissioning itself can be characterised as a complex, dynamic system, open to external influences, including the political and intellectual zeitgeist. We can theorise that the methods specified in commissioned evaluations evolve from the behaviour of actors on both 'sides' of competitive tendering, interacting with a rules-based procurement process (e.g. derived from competition law and public procurement rules). Behaviours are further influenced and shaped by the contexts in which the actors are operating.

What were the aims and approach of the research?

The research sought ultimately to identify the practical steps that could be taken by commissioners and procurement officials to foster the take-up of complexity-appropriate and other less conventional evaluation methods. To identify these steps, it aimed to generate qualitative insight on the following questions:

- What influences the adoption of new evaluation approaches and methods[2] in environmental policy fields?
- What opportunities are there to enhance the take-up of complexity-appropriate evaluation methods?

To do this, the study focused on individuals who are directly involved in competitive tendering – in developing tender specifications and preparing bids. This decision-making nexus is where the influences on individual behaviour, the wider evaluation context, and constraints from the rules and procedures of competitive tendering are all crystallised in the methods that are eventually selected by the procurement process. Specifically, the interviews covered the following themes to investigate the main aims of the study outlined above:

- Why and how are evaluation studies commissioned?
- How does the wider policy and professional context influence evaluation commissioning?
- Is there an appetite for complexity-appropriate approaches and methods in policy evaluation?
- How are evaluation methods selected by commissioners and contractors?
- What are the barriers to new and complexity-appropriate methods?
- How could evaluation commissioning support the uptake of complexity-appropriate methods?

How was the evidence collected?

The research questions were explored through in-depth interviews with nine commissioners and ten contractors who are directly involved in evaluation tendering processes, mainly in, or for, Defra (the UK Department for Environment, Food and Rural Affairs). To ensure consistency in the contexts and procurement processes being explored, the study was restricted to one UK government department.

1 | According to CECAN: "Complexity science is an approach to understanding the world which embraces the fact that it is made up of many diverse components, which interact in adaptive and nonlinear ways within 'complex systems'". See more in the CECAN Manifesto (2018) at https://www.cecan.ac.uk/resources.

2 | "Approach" and "methods" were often used interchangeably by interviewees so no clear distinction is made in this paper. The term 'complexity appropriate methods' is used to encompass both evaluation theories and the application of methods informed by complexity science.

'Commissioners' were senior individuals who are involved in devising evaluation specifications and appointing contractors; 'contractors' were senior individuals responsible for leading bids for Defra evaluations in the last five years. As client and contractor on the same evaluation project, some interviewees were able to give contrasting perspectives on specific examples. Interviews took place in June and July 2018.

It is important to note that the findings reflect the views of the individuals interviewed and do not represent the views of Defra or the contractors' organisations.

The interviews provided good coverage of Defra's different operating areas[3] but limitations need to be acknowledged. The sample was intentionally restricted (for resource reasons) to those creating or responding to tender specifications so it did not cover procurement officials or higher-level budget holders (e.g. deputy directors in the Department), who might have different perspectives. Similarly, aspects of procurement practice may differ in other departments. Further research to test these findings more widely would be worthwhile.

2 | Why and how are evaluations commissioned?

What is the purpose of commissioned evaluation studies?

There is always an internal client and funder for commissioned policy evaluations. The direct commissioner places high importance on meeting the needs of those clients.

The evidence needs of internal policy clients vary according to features such as the political profile and sensitivity of the policy[4] being evaluated, as well as the actual design and mechanisms of the policy. Budgets are typically negotiated with these clients well in advance of the development of a tender, which may constrain a commissioner's options when they come to specify an evaluation methodology. Commissioners and contractors said that their clients' evaluation priorities are (broadly in descending frequency of mention):

- Impact and value for money (including some evaluations that are required to follow standard EU evaluation formats)
- Understanding process and effectiveness
- Learning from pilots or trials to inform policy roll-outs or future development
- Evidence of regulatory compliance

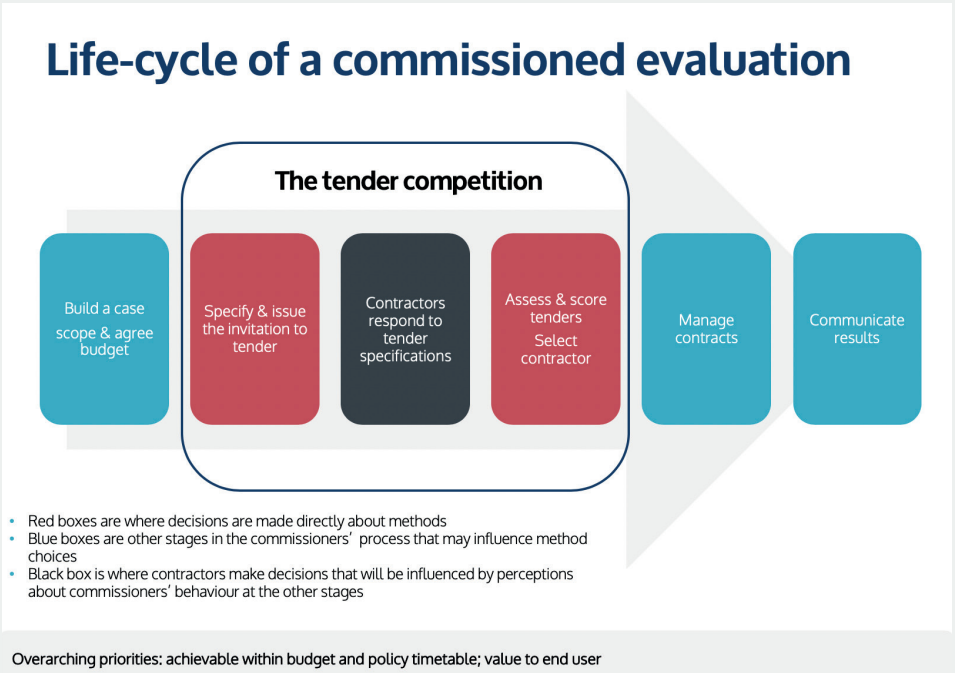

Figure 1: simplified description of the different stages in the lifecycle of an externally commissioned evaluation.

3 | All policy interviewees work in an area that concerns interactions between human and natural systems in different environmental and climate change fields. Contractors covered these and sometimes other policy fields in addition.
4 | Most often, interviewees talked about "policy" evaluation in terms of the outcomes and impacts of policy delivery programmes. Policy appraisal, when policy is being developed, is more usually done by in-house by civil servants.

Common throughout was a strong feeling from those commissioning evaluations that they are under continual pressure to justify how any evaluation spend is worthwhile, and that they are constantly mindful of *"doing more with less"*. That phrase was often cited as a key principle of the austerity drive of recent years in UK government. As we shall see later, those feelings and perceptions help shape appetites for risk taking with respect to 'new'[5] methods.

How are evaluations procured?

For readers not familiar with evaluation procurement, the following is a simplified description of how the process works, to provide background context for the later findings. It also illustrates where evaluation tendering fits within the whole lifecycle of a commissioned evaluation, illustrating upstream and downstream considerations that commissioners and contractors will have in mind during the tender process.

Key points in the commissioning process where decisions are made directly about methods are: the formulation of a tender specification, the contractor's response, and the tender assessment process to select a winning contractor (Figure 1). Individual decision-making at those points will be influenced by individuals' wider contexts and the rules of the competitive tendering process.
On the commissioner side, individuals from (at least) three different professional backgrounds will typically be involved in the process, who have different priorities and interests in the evaluation. Normally they represent:

- **social science** – typically the person who will manage the evaluation contractor and who is a specialist in research methods, though not necessarily evaluation
- **policy** – typically the policy or programme 'owner' and user of the evaluation findings, whose focus will tend to be on what the evaluation will tell them and whether it will deliver on time, as long as they are assured the methodology is robust
- **a procurement representative** – who will typically be concerned with ensuring a commercially fair competition, assuring the financial viability of contractors, and the value for money of the proposed approaches and outputs.

Interviewees most commonly described the following process. To ensure a fair competition, contractors' bids will be scored by the assessors (typically the above individuals) against an assessment framework agreed with Procurement during the tender specification process. Scores for different aspects of the tender proposal will be given different weights. Each tender will be given an overall score, typically combined from: a technical quality score (methodology, appropriate outputs and contractor experience); a project management score (e.g. including risk management); and a financial score.

Crucially, the methodology selected by the commissioning process will be the one that has scored highly across the various scoring criteria, including lowest price. As we shall see later, the weighting between technical and financial scores is an important influence on contractors' decision-making about methods.

3 | Influence of the commissioning context on choosing methods

Before examining how evaluation methods come to be selected in the tendering process, we look here at some of the important wider influences on decision-making that interviewees talked about. That includes their perceptions of how the policy evaluation climate and standards are evolving, and the pressures they have in mind when commissioning evaluations.

Recent changes in the evaluation commissioning context

Interviewees described two recent trends that were sometimes felt to be pulling in opposite directions. (Figure 2).

In response to criticism of weak practice[ii], there has been a general drive to raise evaluation standards across government. This has resulted in new specialist posts being created, and a new central evaluation team for the department had just been set up (June 2018), which was widely welcomed. There was strong support for raising standards.

5 | 'New' is used throughout to describe methods that are new in an innovation sense - either new to the world or new in this application or context.

At the same time, many reported a squeeze on research and evaluation budgets. Where evaluation budget is limited or under pressure, interviewees warned there is a risk of overly narrowing the focus of evaluation onto quantitative impact and value for money metrics (e.g. as defined in HM Treasury Green Book and Magenta guidance). Making a case for methods that are either unproven in these policy contexts, or that cannot provide a simple answer to the question "did it work?", can be challenging.

On the other hand, *"doing more with less"* was a positive for some interviewees: they said it can encourage adaptive and agile policy making based on a test-learn-adapt philosophy, which might encourage the use of new evaluation approaches.

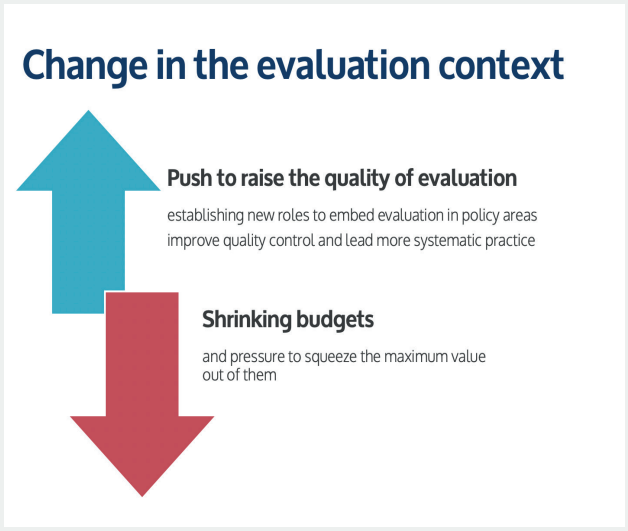

Figure 2: Twin pressures that are driving changes in the evaluation commissioning context

Moreover, not all interviewees were equally affected by budget pressures. Some saw increased resource for evaluation and research in policy areas affected by Brexit; while others perceived that fewer, but larger, evaluations are being commissioned generally, which has created opportunities for multi-method evaluations, including theory-based approaches – but more so in other departments.

Since this research was conducted, and the set-up of the central evaluation function, there have been further developments in evaluation in Defra, including a Complexity Evaluation Framework developed in conjunction with CECAN[iii].

How do operational challenges faced by evaluation commissioners influence what they commission?

Related to those overarching twin pressures on evaluation commissioners, interviewees mentioned a range of connected challenges that are relevant to the choices that can be made about evaluation methods – in addition to any technical considerations about the merits of complexity-appropriate methods. These six challenges emerged from the interviews (the first two relate mainly to commissioners; the rest to commissioners and contractors):

1. How embedded evaluation is in a policy area

Different policy areas appeared to be at different stages of implementing the evaluation improvement agenda, which can constrain what it is realistic to commission in different parts of the department. The stages can be characterised as:

- **Creating demand** – needing to explain the purpose and value of evaluation, introduce basic evaluation knowledge and build skills for delivering core methods (e.g. as defined in the Magenta Book[iv]). A commissioner operating at this level described complexity-appropriate methods as *"a step too far".*
- **Embedding & demonstrating success** - where the case for evaluation and the standard methods are broadly understood, and where past evaluations have helped to improve current policies or secure more funding for delivery. Gaps in knowledge remain but there is openness to alternative approaches if there is good evidence of how they can add value.
- **Improvement & expanding capabilities** – where there has been an uplift in evaluation competence in social science and there is an advanced level of knowledge about a range of approaches; and where thinking is moving on to explore a broader 'evaluation toolkit', including realist, theory-based and/or complexity-appropriate approaches, for example.

2. Whether evaluation is built into policy delivery from the outset or is 'bolted on' later

Commissioners (and some contractors) said that evaluation is often an afterthought, conceived after a policy has been designed and, not infrequently, after delivery has started. This limits methodological options

significantly. Those options may end up being driven by the practicality of what data are readily available or can be collected within the timescale and budget allocated, rather than what is the optimum way to evaluate impact and effectiveness, or to enhance policy understanding. Past reliance on qualitative case study approaches may be one manifestation of this particular challenge.

3. How the budget for a specific evaluation is determined and administered

Limitations flagged here were:

- **Being able to secure enough evaluation budget from internal funders for the most appropriate methods.** Evaluation functions can feel they wield less clout than delivery teams or policy decision-makers where there is competition for budget within an overall fixed amount for the policy programme.
- **Budget approval being linked to a specific methodology:**
  - Funders may have their own preferences which might not align with the research commissioner's view of the optimal method, for example because of their perception of what counts as quality in evidence (as in point 4 below);
  - Or, a methodology agreed as part of an original funding proposal might be overtaken by fast-moving policy and changes to the programme by the time an evaluation is commissioned, but the lengthy approvals process means that it is not feasible to renegotiate the budget or approach.
- **Single-year research budgets** – *"what can be delivered within the timescale and budget"* is often the deciding factor when commissioners and contractors make a choice about methods. Some felt they were bound to commission projects that could be completed within a single financial year, because of annual budget allocations. That can favour proven approaches and avoiding uncertainty (e.g. where data collection is complicated or in methodologies designed for emergent causation). Others (including contractors) thought that can be a constraint but pointed to positive examples of large budget, multi-year evaluations in high profile policy areas.

4. The prevailing evidence culture and what is considered to be 'high quality'

The choices made about methods by commissioners and contractors are constrained by dominant evidence cultures, which largely define what is acceptable to evaluation clients (internal and external). While evaluation practitioners play a part in the evolution of evidence cultures, theirs may not be the most powerful influence on what is accepted as high quality in policy evaluation at any given time.

Demand from evaluation funders and clients for quantitative, counterfactual and experimental approaches (such as Randomised Control Trials) has grown in recent years and is now said to dominate what is accepted as 'quality evidence'. Qualitative approaches and methods (e.g. case studies) have fallen out of favour and may be considered weak or not providing clear answers.

This trend reflects the drive to raise evaluation standards and the influence of the government's What Works Centres[v] and the Behavioural Insights Team[vi].

Commissioners and contractors welcomed the drive to raise standards but generally wanted a more plural evidence culture, favouring a pragmatic blend of approaches from different evidence traditions. Concerns about the dominance of experimental and other counterfactual methods revolved around:

- **Feasibility of their execution in some policy settings** – including issues relating to identifying multiple beneficiaries and outcomes, complexity in policy mechanisms, and securing data of high enough quality to support RCT-type approaches within available budgets;
- **Producing over-simplistic answers** on whether and how complex policies have worked, that are not helpful for improving policy or that risk flawed and failed roll-outs.

Some government departments are *"moving on from the rabid RCT years"* (as one commissioner put it) to embrace a broader range of evaluation methodologies, including theory-based approaches and related methods[vii].

5. The expectations of evaluation stakeholders

Being able to fully meet the interests of powerful internal and external stakeholders is a high priority for the direct commissioners. They will have to communicate and justify the results of any evaluation, often to stakeholders beyond their immediate policy client and evaluation funder, including Minsters. Especially in the context of squeezed evaluation budgets, scrutiny over how well evaluation budget has been used is always at the back of commissioners' minds.

While there has been a general drive towards using more sophisticated quantitative methods in policy evaluation, influential stakeholders interpret the prevailing evidence norms in hugely different ways. As one interviewee said: *"the variation in what they [senior managers] understood as evidence in evaluation was colossal … their professional frames of reference were different…"* so that what stakeholders will accept as legitimate evidence may differ with the individual.

Two issues were repeatedly flagged in relation to meeting stakeholder expectations:

- **Multiple stakeholder interests resulting in too many research questions** – some felt *"doing more with less"* has sometimes led to evaluations becoming *"overloaded"* with *"disparate"* research questions that were too complex (see point 6) for a methodology that would fit the size of the evaluation budget.
- **Simple answers** – evaluation findings need to *"land"* with policy and ministers to have any impact. This might mean needing to produce a quotable statistic on impact, or using the *"right"* kind of method in the eyes of the stakeholders even if data collection is not achievable in the delivery context, or even delivering the *"right kind"* of answer[viii]. Interviewees who were more familiar with complexity-appropriate methods deliberated on whether and how those methods might meet this specific challenge (see "Barriers" below).

6. Complexity in the policies being evaluated

Commissioners are finding that 'standard' counterfactual evaluation methods are not suited to the scale or scope of the questions they are being asked in many cases. This relates especially to the changing context in which they are working and the currency of agile, or adaptive, policy making noted earlier, and "doing more with less". Aspects of complexity in the policies they are evaluating include:

- **Long term, and potentially iterative, outcomes** – including time for building the capacity of local projects and programmes before they can engage effectively with beneficiaries, and behaviour change as an incremental process rather than a one-off response event;
- **Interactions between human and natural systems** – which are inherently complex, and where impacts may only be detectable over the long term and may not be the ones predicted;
- **Devolution of delivery with multiple approaches** – where policy sets the required outcomes and multiple local delivery projects are each left to devise their own approaches, mechanisms and boundaries for the intervention;
- **Area-based programmes** - involving multiple and inter-related target outcomes, multiple delivery partners and multiple beneficiary types;
- **Boundary definition issues** – including geographical boundaries for policies concerned with natural environment outcomes; and programme boundaries where there are multiple agents with differing interests and contributions;
- **Multiple causal paths and combinations of causal factors** – where a mechanism may produce different outcomes in different contexts and with different beneficiaries, or where there are different routes to achieving the same outcome;
- **Variable speed of response** – by beneficiaries in different settings, or subject to different internal or external constraints.
- **The policy landscape in which the policy is being developed** – with multiple stakeholder interests and their operating contexts, as described above.

While those involved in policy evaluation are wrestling with these recognisable features of complex systems[ix] in the settings policies intervene in, those interviewees who are familiar with complexity science stressed that policy and commissioners are largely not making conscious reference to academic concepts when they talk about "complexity".

- Quite often they mean 'complicated' rather than 'complex' and are concerned with how complexity manifests in programmes and evaluations (e.g. aspects outlined above) rather than complex systems as an organising concept for an evaluation approach.
- Some warned that the academic language of complexity would be off-putting to policy and evaluation funders, especially in those areas where evaluation is not embedded.

4 | Evaluation commissioning and complexity-appropriate methods

Is there an appetite for complexity-informed approaches in policy evaluation?

Commissioners' views on this question are summed up by the quote in the box. If complexity-appropriate methods can be shown to give evaluation users the answers they need, then the methods will be used.

> "...complexity is no different in that you need to make sure that the outputs are genuinely useful otherwise it is a piece of academic work that doesn't land in practice...."

Some commissioners have a worry that complexity-appropriate methods either are, or may appear to be, too *"academic"*. The professional risk to the commissioner is that findings are viewed by their internal clients as interesting but not directly usable, whether that is for accountability or to steer policy improvement. While that might not be the case when methods are applied in practice, it is a perceptual barrier that needs to be overcome.

Contractors said they are largely driven by what the market demands, even though they may promote their own intellectual interest in new methods with long established clients. Some had started to see tenders calling for such methods but mainly in other government departments.

Commissioners in this department were not yet seeing that kind of demand from policy clients, either for complexity-informed approaches or other alternatives to 'standard' evaluation methods. Some indeed felt their colleagues and internal clients are not ready for the mind-shift that complexity-informed approaches would imply[6].

The exception was those who have been involved with CECAN who felt confident in their own position and knowledge to advocate for complexity-appropriate approaches, or who had identified a low-risk project to test whether a specific method could deliver a better evaluation outcome.

Elements that would need to be in a policy-focused business case for complexity-appropriate methods, based on this research, are drawn together at the end of this paper.

How are evaluation methods selected by commissioners and contractors?

Commissioners

A commissioner's preferred methodology for an evaluation is set out in the tender specification (or 'invitation to tender') which contractors then respond to in competition with each other.

Tender specifications are typically the outcome of a back and forth dialogue between the lead research manager, policy and procurement. Time is commonly a constraint on finalising a specification: it is often short, which interviewees said will tend to focus attention on well-understood methods. Where evaluation is 'bolted on' at a late stage some approaches and methods also have to be ruled out because they are not feasible within the delivery mechanism, budget or timescale.

More rarely, commissioners said they were able to undertake early scoping studies (either internal or externally commissioned) which enabled a wider range of methods to be considered before one was chosen for the tender specification.

Most often commissioners are striking a compromise between the methods that are technically best-suited to answer the evaluation research questions, what is achievable with the time and resources available, and their policy clients' and funders' preferences. Commissioners cited these main influences on the methods they propose in tender specifications:

- Appropriateness to the evaluation research questions
- The approach that was agreed with the funder
- Constraints of budget and timescale
- The limits of personal knowledge of methods
- Guidance in the Magenta and Green Books
- Personal confidence in securing a useful and defendable result (i.e. not career damaging)
- Confidence that methods can be scored fairly and effectively in the tender assessment process so that the tender exercise will succeed (see under 'Barriers' below)

6 | As noted earlier, a Complexity Evaluation Framework has since been developed by Defra and CECAN.

Contractors

An important aim for commissioners is to issue a tender invitation that will attract enough bids to ensure an effective competition. Contractors will not always respond if they feel an evaluation is unrealistic (e.g. scope, timescale, budget), too risky, poor quality, or is not intellectually rewarding (more than half the contractors, especially small ones, said this). The latter is not only a personal preference but is also related to professional development, corporate identity and profile in their markets.

Most contractors perceive creativity as an important contributor to their competitive advantage so are often looking for ways to challenge commissioners' specifications and/or to offer innovative approaches. However, they need to be careful how far they can push the specification to add value while remaining compliant with the competition rules (e.g. in terms of meeting the tender scoring criteria). It was very clear from the interviews that contractors are frequently second-guessing what commissioners really want and how open they are to innovation, not least because the usual tendering process rules out meaningful dialogue between the parties. The risk of guessing wrongly and not winning the commission can favour conservative choices of method.

How interest in creativity and innovation is signalled in the commissioning process

Both commissioners and contractors talked extensively about the choice that commissioners need to make on whether to put out a tender invitation based either on:

- a tight specification - one where the approach, methods and outputs are specified in detail; or
- a more open specification - where required outcomes are clearly defined but there is scope for contractors to interpret the best ways to achieve those

Both sides can see advantages and disadvantages of each approach (Figure 3): but there was consensus that open specifications leave more room for new methods to emerge, unless the commissioner has explicitly specified a 'new' approach in a tightly defined tender specification.

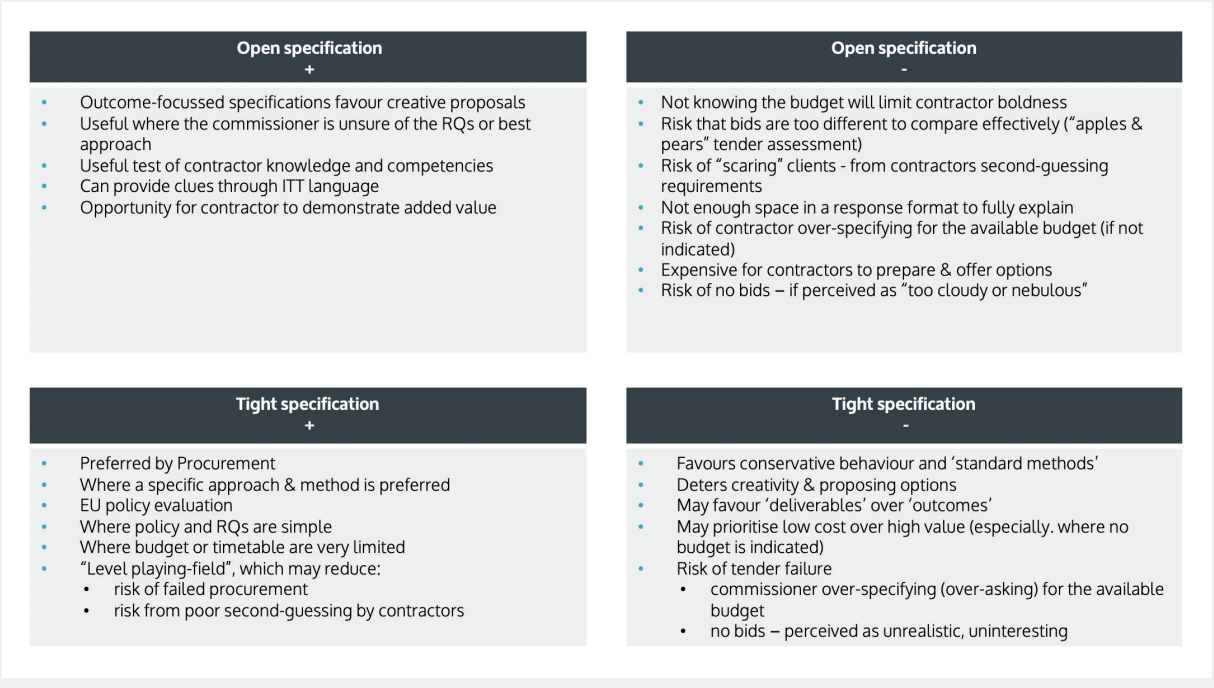

Figure 3: Characteristics of tender specifications that were reported to favour (+) creativity and innovative methods in tender responses or to act as barriers (-)

While open specifications may encourage creativity from contractors, a major deterrent to specifying innovative methods is there being no budget guideline in the tender invitation (which is normal practice in the department according to interviewees). Not knowing what the price 'floor' is that competitors could propose will tend to deter more sophisticated but more expensive methods because contractors will not be able to

take an informed view on how much ground they would have to make up on a superior technical score, and whether that is realistic. This factor alone drives conservative behaviour by contractors who will tend to focus on the minimum way to achieve the requirement at the lowest price, rather than taking a risk on a higher cost, more optimal methodology.

What are the barriers in commissioning processes to adopting innovative methods?

Interviewees were asked what barriers there are if they want to propose innovative evaluation methods, in either a tender specification or response, including methods suited to evaluating complex policy.

All of the barriers that interviewees mentioned are relevant to the take-up of complexity-appropriate methods, as well as methods innovation in general. On top of those barriers are some more specific ones that arise from the intrinsic characteristics of complexity-appropriate methods. The general innovation barriers are outlined first then the barriers relating specifically to complexity-appropriate methods in the subsequent section.

Barriers to commissioning innovative methods reflect interactions between:

- Individual actors' attitudes and behaviours around risk
- Aspects of the commissioning process
- Feedback loops with the wider policy and operating context

1. Attitudes to risk tend to favour established methods and deter innovation

In talking about barriers to the adoption of innovative evaluation methods, it was apparent that individuals' attitudes and behaviours around risk were one of the most important. Those attitudes and behaviours are influenced dynamically by the individual's past experiences of commissioned evaluations, their relationship with funders and procurement teams, and their own wider professional context. Some said that greater pressure on research budgets had heightened their and colleagues' risk aversion in evaluation commissioning. Risk barriers for commissioners include:

- Lack of appetite for experimentation in the context of tight budgets and timelines
- Difficulty of assessing the likely cost and value for money of new methods, and therefore not being confident to specify it in a tender against an already agreed budget
- Lack of knowledge about new methods and concerns about their own competence to manage the contractor and to quality assure the evaluation outputs

For both commissioners and contractors:

- Fear of the unknown in terms of what outcomes new approaches will deliver and being able to explain and justify methods to clients (internal or external)
- Difficulty in assessing the delivery risk of unknown methods
  - Risk of evaluation failure, and resulting risk to professional standing
  - Resource risk of applying methods to a new policy and/or delivery context and 'learning on the job', with consensus that this investment risk is borne largely by contractors

Contractors said they would be led by the market. A key risk for them is lack of client demand to balance against the scale of investment that would be needed to upskill their teams in 'innovative' methods. A few described examples where there had been an element of risk sharing with the government evaluation client, including flexibility to re-profile resources and deliverables at key stages during the evaluation of a very complex programme or where they were applying a novel method.

Mitigating risks of the unknown can also happen where the client and contractor have a long-term working relationship. It can enable shared learning over several projects and build mutual trust in finding solutions to cope with the unexpected – but this type of co-productive learning relationship is specifically discouraged by competitive tendering. Procurement models are designed for an arms-length purchase-fulfilment relationship, and not for delivering shared goals and learning benefits, or for sharing risk.

2. Aspects of the commissioning process that deter methods innovation

While most interviewees make the best of the commissioning system as it currently operates, some questioned whether it is fully fit-for-purpose for evaluation commissioning, and for social research more generally.

Research procurement in government has developed from systems designed for procuring physical products, for which a quality standard and required quantities can be specified and contractors compete on price and ability to meet or exceed the quality standard for the best market price. All parties involved in the competitive process need to be kept separate to prevent collusion and thus ensure that the procurement authority achieves an economically efficient price and best value for money. Innovation risk and investment is borne by contractors striving to optimize their competitiveness and profitability.

Those underlying procurement principles become apparent in evaluation commissioning in procurement preferences for tight specifications. As outlined earlier, tight specifications tend to favour measurable outputs in tender scoring and contract Key Performance Indictors (what several referred to as "tick box" practice), inflexibility for contract variation, and in preventing meaningful dialogue between commissioners and the market at the scoping stage. Criticisms levelled by interviewees at this approach included:

- Methodologies geared to quantity rather than quality
- Unhelpful or failed evaluations where more evolutionary or recursive approaches (whether designed in from the start, such as theory-based evaluation, or in response to learning during delivery) were discouraged by rules and/or behaviours at different stages of the procurement process and contract delivery.

Within the commissioning process, specific barriers and constraints can be experienced at any or all of these stages according to interviewees:

- **Upstream from the tender specification, especially:**
  - Being able to sell innovative methods to internal clients
  - The prevailing evidence culture that favours certain types of approach, method and metrics
  - Procurement rules that restrict meaningful dialogue between commissioners and contractors before a tender exercise - which limits open exchange of knowledge and ideas on approaches best suited to meeting the commissioner's evaluation need
- **In the tender specification and contractor response:**
  - Not providing a budget indication, which encourages conservative choices about methods by contractors
  - Procurement officials and some social science commissioners preferring 'tight' specifications (see discussion above) based on 'standard' evaluation methods
  - Procurement and researchers having different objectives for the tender exercise and different understandings of the implications of methods choices, including to the value and usability of the findings
  - Tender response formats and IT platforms that make it difficult to provide a coherent account of a new methodology in a tender response, including:
    - Specific formatting limitations (e.g. no tables or diagrams, splitting response forms into too many sections to enable a coherent account of the method)
    - Space and character-count restrictions for explaining the methodology and needing to focus limited space on outputs and 'deliverables' to maximise tender score
    - No response sections to propose creative alternatives or 'added extras' and the risk to contractors of being penalised in the price scoring for including extra elements in the core methodology
- **At the tender assessment and contract award stage:**
  - Tender scoring with a high percentage of the total marks awarded to price (more than 30% or 50% were mentioned), where offering an alternative will make insufficient difference to a contractor's positive technical scores and could undermine their price score if adding methodological value also adds resource cost
  - Specifications where scoring and Key Performance Indicators (KPIs) for delivery are tied heavily to measurable outputs (or 'deliverables': for example, a given number of meetings, workshops, or a survey, as opposed to an evaluation process (e.g. quality of stakeholder involvement) or the value of outcomes
  - Splitting a tender into separate sections for scoring, including criticism that value for money formulae applied by procurement officials may conflate cost with value and unfairly penalise aspects of methodologies that are higher unit-cost but also higher unit-value to the evaluation
- **Post-tender award - contract management and compliance:**
  - The contract process can work well for evaluations where policy and research questions are straightforward, specifications can be tightly defined, and results are time-critical for policy

- However, where flexibility is needed there is often limited scope to vary contracts once they are awarded, without the contract being at risk of non-compliance and needing to be re-tendered, including inflexibility around KPIs, deliverables and timelines

Where evaluations are more complicated or have features of complexity, however, this inflexibility is another factor that tends to favour conservative selection of known and trusted methods on both sides of the commissioning divide. It equally makes it difficult to specify and deliver contracts that involve significant elements of learning, feedback, emergence and methodological development during delivery – as would be the case with complexity-appropriate evaluation.

"Complexity-appropriate evaluation is a rigorous and holistic approach that makes evaluators' work both easier and more effective. It is an approach that:

- emphasises adapting to emerging findings;
- mandates iterative cycles of design, data collection and learning;
- engages a wide spectrum of stakeholders at all stages;
- embraces the full complexity of the policy and context being evaluated
- assumes we can only steer complex systems, rather than control them fully."

Policy Evaluation for a Complex World, CECAN, February 2018

Figure 4: CECAN definition of complexity-appropriate evaluation

Are there intrinsic features of complexity-appropriate approaches and methods that make them difficult to commission?

The question of what we mean by 'complexity-appropriate' evaluation and methods is relevant here.

CECAN defines a complexity-appropriate evaluation as one that is framed in the terms of complex systems and has the characteristics outlined in Figure 4[x]. A complexity-appropriate method is therefore one that has the ability in some form to capture and support improved understanding of one (or ideally more) of the characteristics of complex systems (i.e. path dependency, feedbacks, radically open systems, emergence etc). Methods can (and have) been re-purposed from other settings to use within such a complexity-informed evaluation framework.

As we saw earlier, interviewees had different understandings of complexity and this is reflected in many of them being uncertain whether complexity-appropriate methods are intrinsically different or less easy to adopt than any other novel method.

Certainly not all of them were thinking in the terms of the CECAN model of complexity-appropriate evaluation; and only half were familiar with the types of complexity-appropriate methods being supported by CECAN (more often theory-based and realist approaches with Qualitative Comparative Analysis (QCA) or process tracing than novel modelling-based approaches[ix]).

A few individuals were slightly dismissive of what they called "*CECAN language*", by which they meant descriptions that made methods sound overly difficult or academic (e.g. "*fuzzy cognitive mapping*"). Language matters not only to how receptive individuals are but also, some suggested, in being able to 'sell' new methods to their internal or external clients.

Reflecting their different levels of understanding and knowledge about complexity-appropriate evaluation most interviewees indicated a qualified "maybe" when asked directly if there are any intrinsic features of the methods that would be a barrier for commissioners.

In addition to their general concerns about innovation risk outlined above, concerns tended to revolve around the following practical features:

- How to achieve flexibility in procurement and contracts to accommodate unpredictability and emergence
- How to enable collaborative working, which is essential to some of the approaches and methods
- How to accommodate multi-stakeholder perspectives and involvement
- Concerns about cost and timeliness
- Concern about the usability of findings

Figure 5

Potential barriers arising from intrinsic features of complexity-appropriate methods cited by interviewees

Pre-specification Concerns about policy timeframes, costs and the usability of findings	Specification and tender assessment Scepticism that a 'rigid' procurement process can cope with emergence and fairly assess the value of complexity-appropriate approaches	Contract management Concerns about enabling flexibility - to enable changes in scope and activity, collaborative working and shared learning
Perception that the evolutionary nature of some complexity-informed approaches will not be able to deliver "answers" within tight policy timeframes or offer guarantees on when results will be available. Concern about the usability of findings - whether complexity-appropriate methods can: • deal adequately with questions of attribution • deliver clear-cut conclusions, or whether outcomes will be too "woolly" or nuanced for clients (internal or external) - including in multi-stakeholder evaluations • related to both points, whether the methods offer an equivalent to counterfactual approaches (e.g. where funders need 'simple' answers) Perceptions or experience that some methods are intrinsically more expensive and/or need a longer timeframe than well-known alternatives. • For example, multi-stakeholder and collaborative approaches that involve senior individuals in extensive and ongoing processes - as compared to 'cheap' approaches such as online surveys.	Perception that some complexity-appropriate approaches are intrinsically at odds with a rigid selection process that requires methodologies to be turned into: • a set of tangible outputs and time-bound milestones • to be delivered in a linear progression • and against which contractor performance will be monitored A belief that the flexibility in 'deliverables' that may be required in complexity-appropriate methods would typically be discouraged and/or penalised in tender responses. Concern about how the process evaluates relative 'value for money' of conventional and alternative methodologies, notably in comparing tangible and intangible value (e.g. surveys versus participatory and co-creation processes). A belief that procurement officials may not understand how value is created in complexity-appropriate approaches and therefore (unintentionally) create bias in the tender scoring process in favour of traditional methods.	Belief that it is essential to create space for revising the scope, budget and the timelines of 'deliverables' in response to emergent findings in complexity-appropriate evaluation - but is difficult in existing contract management approaches. The scale and nature of flexibility can't be anticipated at tender stage - some say that very large changes would render contracts void, forcing a re-tender exercise. Where changes can be accommodated, contractors are seen to bear most of the risk of extra cost - which is a deterrent to proposing the methods. Perception that procurement rules actively deter collaboration: they set up contractors to be separate and independent operators, and commissioners to be enforcers. Collaborative working would require additional resource for both commissioners and contractors (which itself could be a disadvantage in tender scoring).

A more fundamental challenge, mentioned by a few, would be a need to shift evaluation funders' and users' mindsets from a linear to systems-based way of thinking about policy delivery and accountabilty. The implications of that higher-level challenge are discussed later.

Returning to the practical challenges, the Figure above summarises the various hurdles identified by interviewees, organised according to the different stages of the commissioning chain where barriers might occur. These are often special cases of the general barriers to methods innovation identified earlier.

How would commissioners respond to tenders that propose complexity-appropriate methods?

Against the background of their views on barriers to new methods, interviewees were asked how they thought commissioners would respond if a contractor proposed a complexity-appropriate approach and methods.

Responses were a mix of anxiety about risks of the unknown and excitement that it could add something new and valuable to the evaluation 'toolbox'. The different evaluation and policy contexts in which individuals work are clearly an influence on their openness to trialling such methods. Those who are open to the idea need to be able to sell unfamiliar methods to their internal clients and to provide guarantees that it will deliver the answers and learning that policy requires. To do that they would need:

- Proof that it had been used successfully, in policy as well as academic examples
- Evidence that it has worked in similar policy and evaluation contexts
- Proof that the contractor has the skills and knowledge to deliver it
- To do their own background research to verify other examples and to understand the capabilities of the method (rather than the detail of how it works)
- A description of the method in everyday language, to help them understand its capabilities and communicate the benefits to internal clients and funders, who are not experts in research methods

Commissioners would normally be looking for those features in any tender response but there was a sense that bids proposing novel methods would need to achieve higher thresholds to achieve winning scores, given the evident nervousness of many commissioners about being able to provide guarantees to their internal clients. Some were also sceptical that the current approach to tender assessment could be made to work in a way that would not discriminate against complexity-appropriate methods and may need a radical overhaul instead.

5 | How could evaluation commissioning support the uptake of complexity appropriate methods?

The evidence here is drawn together from responses to the following questions that were asked of all interviewees at different points in the interview:

- How much flexibility is there for evaluators to be creative or innovative in their bids?
- How could procurement barriers be addressed by commissioners?
- What questions would commissioners have about complexity-sensitive methods? How could those be addressed?

An enabling environment for innovation in evaluation methods

Thinking first about the general environment in which the commissioners (mainly social scientists) are operating, a number of improvements and enablers were suggested to support methods innovation generally:

- Access to expertise in evaluation methods and new developments in the field, including:
  - A call-off panel of experts[7] or capacity to commission external scoping studies
  - Upskilling in evaluation - both ring-fenced time for their own learning and on-going evaluation training for policy teams
- Embedding evaluation earlier in policy programmes and considering evaluability in the design of delivery mechanisms, which would allow the applicability of different types of evaluation to be considered
- Champions of novel methods and "constituencies" of knowledge, which are seen as having been central to recent methodological developments but have depended on individuals taking risks and driving innovation rather than it being a strategic approach across or within departments[8]
- De-risking the application of new methods by starting small and trialling them in new policy areas where there is less evaluation history and barriers from embedded methodological preferences: for example, Brexit-related policy or individual project opportunities where a new approach could be tried out at low cost and risk because it is not a high-profile policy.

Changes within the existing procurement framework and process

Turning to procurement directly, many commissioners and contractors think there is scope to support innovation and the uptake of complexity-appropriate methods within the present commissioning system.

Suggestions were made for changes at each stage in the existing evaluation commissioning process (Figure 6). While we can describe them individually, many would only be effective enablers of new methods in combination with changes at other points in the process. For example, encouraging collaborative evaluation approaches in bids would also require changes to the way in which contracts and contractors are managed.

1. More routine use of scoping in recognition of complexity in policy and evaluation

Uncertainty around what complexity-appropriate methods would deliver, how they would work, what the risks are, and how much they would cost could, in principle, be reduced by undertaking evaluation scoping studies more often where policies are complex. This was one of the interviewees' most frequent suggestions.

Early – and crucially more open than at present – dialogue with contractors might also help to narrow down scope, which would aid the assessment process so that *"you'd have five apples and five pears rather than twenty-five apples and pears"* to score against each other in response to a specification that encouraged creativity.

Those who wanted this were not hopeful that procurement processes could be reformed to enable the kind of honest dialogue they envisaged. At present, contractors are consulted by commissioners as a group (where this happens) and their clarification questions are shared with all bidders so that they will be cautious about indicating too much of their thoughts and intentions to competitors. Alternatives that interviewees believe are allowable under current rules could include staged procurements, where an early expression of interest (EoI) stage is used to narrow down the number of different ideas being put forward; or holding

7 | Several cited panels that existed at BEIS and DfT; also a hope that the new central evaluation function in the department would fill this role

individual bidder interviews or pitches before or after the tender assessment.

As noted earlier, time for scoping and *"time to think and reflect and talk through the possibilities without kind of ruling anything out too soon"* is however rare and is often impractical because of policy and procurement timelines. Interviewees said scoping can be done where evaluation is embedded right at the start of policy development and gave examples of it happening in new policy areas.

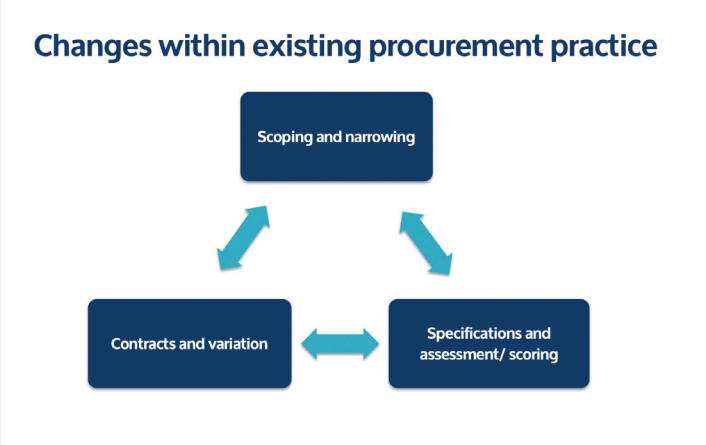

Figure 6: Stages in evaluation commissioning where changes could be made to support the use of complexity appropriate methods

2. Finding ways to accommodate uncertainty and encourage flexibility in tender responses

At the next stage, overcoming barriers that arise from specifications and the evaluation assessment scoring might include:

- Issuing outcome-focused or less prescriptive specifications, with enough information for contractors to make informed decisions about how far to be creative when innovative approaches are explicitly encouraged, crucially including a budget indication
- Developing scoring criteria that encourage and reward:
  - Ways of working, quality of delivery, insight and usability of findings, as well as tangible and easily measurable 'deliverables' (i.e. is more attuned to selecting the best value for money service and research result, as opposed to a tangible output)
  - Contractors who put forward value-added options – but recognising that inviting a compliant and an alternative approach has a cost for contractors in terms of bid preparation, so keeping it proportionate to the size of the commission
  - A post-award scoping stage, where this is essential to the type of approach being proposed
- Avoiding value for money measures based on price per unit of output (e.g. workshops and stakeholder engagements), which can favour contractors who *"promise the moon"* but where the quality and appropriateness of individual outputs may be poor
- Where the precise scale of specific tasks or outputs cannot be fully anticipated, base scoring for these elements on contractor day rates rather than total costs for specific tasks or outputs (this is something that commissioners already do in some cases)
- For individual evaluation commissions that are seeking innovative methodologies, reviewing the appropriateness and usability of standardised response templates: do they enable the contractor to explain the approach fully and the commissioner to understand the benefits and risks of what the contractor is proposing in a joined-up way?

3. More flexible and responsive contract management

Once contracts are awarded[9], the flexibility that many said would be needed to support the use of complexity sensitive methods could include:

8 | Though several who talked about this noted that the situation is changing, albeit still at an early stage in some parts of the department.
9 | These measures would need to be included in the tender specification as they will influence costs and approaches that contractors propose.

- Using measures that are already available to commissioners:
  - Include a post-award scoping stage to finalise methodology, with the extent appropriate to the approach chosen (e.g. for some realist evaluation this could be an extensive part of the project)
  - Include stage gates for review of methodologies, interim outcomes and activities
  - Enable flexible deployment of resource within identified stages of evaluation projects, tied to outcomes rather than outputs
- Careful consideration of KPIs for contract delivery that are appropriate for emergent methodologies, which enable deliverables and milestones to shift where that is justified by ongoing learning in the project rather than a result of poor performance
- A new, more open, agile and collaborative approach to contract management, including:
  - On-going dialogue rather than intermittent reporting against milestones
  - Collaborative learning between commissioner and contractor
  - Responsive resourcing within overall budget envelopes
  - A live, shared, risk register: for example, to manage shifts in deliverables and milestones in a transparent and accountable way
  - Active management by the commissioning research manager and additional management resource to support it than in traditional evaluation commissions[10]; line management that supports flexibility in time use by research managers and prevents overload
  - Development of new research manager competencies that include being comfortable with uncertainty and adaptability under time and delivery pressure;
- Learning about procurement and contract management from other departments that have used non-standard evaluation methods more extensively (including developmental evaluation approaches)

"...accepting that it's going to be emergent and will need regular reviews of the methodology of the data to really understand whether you're getting the right data, using the right methodology."

Contractor

Commissioners, mainly, thought it would not be easy to change the way in which evaluations are commissioned and felt they have little influence as individual users of the procurement system. They can negotiate with procurement officials about the tender assessment criteria during the specification stage but within strictly defined boundaries that allow only limited flexibility. Four notable barriers to change within existing procurement practice were mentioned:

"...in the real world it doesn't work like that. In the real world you never have time...You can conclude and write it down, 'procurers should allow more time', but it just never happens."

Commissioner

- Constraints on individuals' time to influence or engineer change
- Timeliness of evaluation delivery – the multiple factors in the evaluation development chain that tend to result in short time horizons for running the procurement exercise
- Embedded practice and preferences, including risk aversion that arises from multiple causes (as outlined above)
- Lack of shared understanding between researchers and procurement officials, including how decisions about methods will affect evaluation findings and their usability in policy[11]

Higher-level actions to support the take-up of complexity-appropriate evaluation

Some suggested that tweaking procurement would not be enough, however. The most critical interviewees believe the present 'one-size-fits-all' model is inadequate for guaranteeing that the most appropriate methods will be selected. They want procurement to be simpler, quicker and more flexible.

"So, I think a whole different approach to procurement would be needed, and one that makes apples and pears being compared much more feasible."

Commissioner

10 | The same would be needed for contractors but was not mentioned in relation to contract management; their concerns were focused earlier on whether the tender scoring framework would penalize them for adding extra resource if it is essential to a complexity-appropriate method.

11 | But noting this is a one-sided view because procurement officials were not interviewed for this research.

To support the take- up of complexity-appropriate methods, they thought a more radical overhaul would need to happen at a higher level in policy environments to influence the attitudes and behaviours of end-users of evaluations, as well as procurement teams. That could help commissioners to procure the approaches that are most suited to each evaluation, encourage innovation (internally and externally in the contractor community), and enable new styles of contract management that are more collaborative.

> "...there is something to be done there at a level above where the ITT gets issued in terms of actually saying 'We are open to these things, these are legitimate methods that should be proposed that we actively encourage'"
>
> *Contractor*

Strategic change in government tends to be difficult and slow (as some noted) and at the mercy of reversal or revision as individual decision-makers change, so there was little optimism that procurement processes and practice would change in any significant way in the short-term.

In many ways the problem of supporting the take-up of complexity-appropriate methods is a specific case of an innovation problem. We can observe that a few early disruptors, working at lower levels of the system (commissioners and contractors), have begun to advocate and demonstrate the use of non-standard evaluation methods. This is happening in disparate pockets of the department, and in some government departments but not others. Complexity-informed practice is evolving but is at an early stage and is not common at higher decision-making levels of the system.

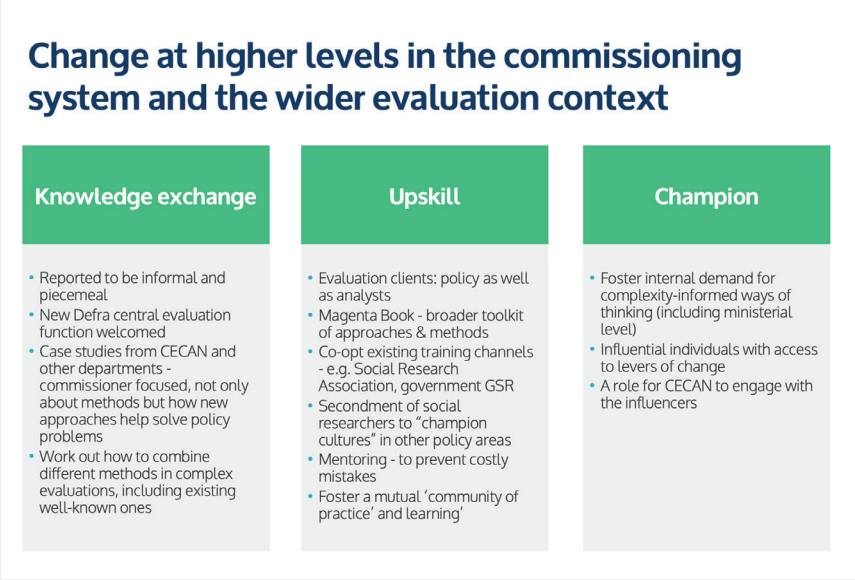

Figure7: Practical suggestions from interviewees to support the take-up of complexity-appropriate methods in policy evaluation

Many suggested that a positive step would be to foster internal demand among research professions and higher-level decision-makers (including ministers), which would help to normalise the use of non-standard and complexity-informed methods. Demand from policy for complexity-informed approaches would help to de-risk decisions to use those methods for individual commissioners or contractors.

That would require (see Figure 7):

- Formal channels for knowledge exchange about complex systems ways of thinking and appropriate evaluation methods
- Upskilling of evaluation commissioners (social scientists, budget holders) and evaluation users
- Internal champions, including influential individuals at higher levels in research-related professions (e.g. chief scientists, social scientists, economists) and evaluation users in government

Making a business case for complexity-appropriate methods

In addition to normalising complexity-informed ways of thinking across evaluation and policy communities, individual commissioners and contractors described various elements that together would make a business case for complexity-appropriate or other innovative methods at the level of an individual evaluation commission. This would also help them to 'sell' such approaches to evaluation funders and end-users.

Interviewees who knew about CECAN case studies thought those should be used to provide the evidence

to answer business case questions. Several commissioners stressed that they need to know less about the detail of *how* a method works and more about *whether* it will work in their context and that it *will deliver* a result that is useful to policy: also, what markers they need to look out for in a tender proposal to be able to commission it effectively.

Figure 8: Questions that commissioners would need to answer in a business case to justify complexity-appropriate evaluation

6 | Implications for evaluation commissioning and promoting complexity-appropriate methods

With respect to evaluation commissioning this research set out to identify in very practical terms:

- What influences the adoption of new evaluation approaches and methods in environmental policy fields?
- What are the opportunities to enhance the take-up of complexity-appropriate evaluation methods?

It has shown that many of the barriers to the take-up of complexity-appropriate methods are related to the novelty and lack of track-record of the methods. This may be a temporary situation which reflects a specific moment in time in an innovation trajectory for new evaluation methods, including complexity-appropriate ones. There were signs from the research that change is happening in localised parts of the government evaluation community: CECAN has been part of that change.

Various ways to foster an environment that supports methods innovation in government research professions were highlighted. Many of these suggestions are not new in research and evaluation communities. An essential component will be to increase the demand from evaluation users for methods innovation. Drive from suppliers and researchers alone will not be enough to shift practices. Encouraging demand will require upskilling of users as well as suppliers, formalised channels for knowledge exchange, development of a community of practice in these methods, and champions in positions of influence in government departments.

Risk is a crucial influence on decisions made about methods during the whole commissioning chain, from securing approval and budget for an evaluation all the way through to communicating the findings to policy users. It includes individuals safeguarding their own positions by adopting conservative behaviours. Supporting the take-up of complexity-appropriate methods will need to include de-risking the use of the methods in a context where they are at an early stage of adoption. This will require advocates such as CECAN producing the evidence to populate the business case identified in this research. It will need to focus as much on demonstrating the appropriateness, comparative value and risks of these methods as the technical detail of how they work.

Underneath, and interacting with, these higher-level barriers are the barriers that arise from the way in which evaluation procurement is currently configured and practised. The research identified changes that might be made at key points in the existing competitive tendering process to make it easier for novel methods generally to compete effectively against traditional evaluation methods:

- Before or while tender specifications are being developed – including evaluation scoping studies and/or measures to enable *meaningful* consultation with suppliers
- Configuration of tender specifications and competition scoring criteria where 'creative' proposals are invited, addressing the barriers in this research that identified drivers of conservative methods choices, including:
  - Scoring frameworks and value for money formulae, paying attention to capturing quality of evaluation outcomes and knowledge creation in the client, as well as quantities and delivery of tangible tasks
  - Providing a budget guideline
  - Treating risk and project management as integral to methodologies and not assessing them separately from the technical approach
- Contract management – more flexibility to enable learning about methods as they are delivered, without risking contract failure through the contractor being unable to meet KPIs that have become redundant (e.g. in long-term, developmental or very complicated evaluations)

All of the above apply equally to supporting the take-up of complexity-appropriate methods. In addition, the research identified features of some complexity-appropriate methods that are especially difficult to accommodate in existing procurement processes, and could require more radical change to procurement rules and practices:

- Flexibility of scope, tasks and resource allocation – to accommodate emergence in the parameters being evaluated and responsive evaluation frameworks and tasks
- Co-productive ways of working – commissioner-contractor relationships based on trust rather than command-and-control, mutual learning, some shared risk taking, and resource to support the active management and ongoing dialogue needed to make it effective

Change here would require a lead from procurement functions in government, to explore alternative commissioning models that enable more collaborative approaches than are possible in conventional competitive tendering and contract management. Models like this may exist[xii] but were outside the scope of this fellowship and would require further research.

Notes

i. See CECAN case studies at https://www.cecan.ac.uk/case-studies
ii. For example: National Audit Office (2013) Evaluation in Government. https://www.nao.org.uk/report/evaluation-government/
iii. Defra Complexity Evaluation Framework https://www.cecan.ac.uk/news/complexity-evaluation-framework
iv. HM Treasury (20110 Magenta Book https://www.gov.uk/government/publications/the-magenta-book
v. Information about the government's What Works Network https://www.gov.uk/guidance/what-works-network
vi. Behavioural Insights Team https://www.bi.team/
vii. CECAN is helping to develop a new annex to the Magenta Book on complexity and complexity-appropriate methods.

viii. Ensuring the independence of evaluations was a concern flagged in the NAO report "Evaluation in Government" and is an issue that is widely mentioned in evaluation guidance (e.g. Stern E for Bond (2015) Impact Evaluation: A Guide for Commissioners and Managers). It is one of the reasons why collaboration and co-creation in evaluation methodologies may be viewed negatively by some commissioners and procurement.

ix. CECAN Fellow, Joanna Boehnert developed a useful visual representation of the key characteristics of complexity https://www.cecan.ac.uk/sites/default/files/2018-06/The%20Visual%20Communication%20of%20Complexity%20-%20May2018%20-%20EcoLabs.pdf

x. CECAN (2018) Policy evaluation for a complex world. https://www.cecan.ac.uk/sites/default/files/2018-02/Cecan-Manifesto%20%2821%20Feb%202018%29.pdf

xi. See the CECAN Resources page for examples of different methods. https://www.cecan.ac.uk/resources

xii. Provisions in updated EU procurement regulations in 2014 may offer scope for more flexible consultation with the market and procedures for situations where services are entirely new. Expert advice would be needed to explore whether these could be used in evaluation procurement. See: Procurement Policy Note: Availability of Procurement Procedures (Decision Tree) https://www.gov.uk/government/publications/procurement-policy-note-1215-availability-of-procurement-procedures-decision-tree and a very accessible guide to the implications of the changes at https://www.bateswells.co.uk/file/the-art-of-the-possible-in-public-procurement-pdf.

www.cecan.ac.uk / cecan@surrey.ac.uk / +44 (0) 1483 682769

The Centre for the Evaluation of Complexity Across the Nexus (CECAN) is a £3m national research centre hosted by the University of Surrey, which brings together a unique coalition of experts to address some of the greatest issues in policy making and evaluation.

July 2019